MULTIPLIER BOOTSTRAP MEETS HIGH-DIMENSIONAL PCA: THE GOOD, THE BAD AND THE MODIFICATION

BY XIUCAI DING[1,a], JIAHUI XIE[2,b] AND LONG YU[3,d] AND WANG ZHOU[2,c]

[1]*Department of Statistics, University of California, Davis,* [a]*xcading@ucdavis.edu*

[2]*Department of Statistics and Data Science, National University of Singapore,* [b] *jiahui.xie@u.nus.edu;* [c]*wangzhou@nus.edu.sg*

[3]*School of Statistics and Management, Shanghai University of Finance and Economics,* [d]*yulong@mail.shufe.edu.cn*

In this paper, we examine the feasibility (i.e., both the good advantages and the bad limitations) and the adaptivity (i.e., the potential for beneficial modifications) of employing multiplier bootstrap to analyze the asymptotic distributions of the largest eigenvalues of potentially spiked high-dimensional sample covariance matrices. Our findings and proposed algorithms demonstrate that multiplier bootstrap remains valid, provided the multipliers are appropriately chosen and the bootstrap procedures are applied multiple times with suitable corrections. First, for non-spiked sample covariance matrices, we propose a novel algorithm to replicate the asymptotic distribution (i.e., the Tracy-Widom law) of their largest eigenvalues. This is achieved by repeatedly bootstrapping the entire sample covariance matrix using carefully designed bounded multipliers that satisfy certain concentration properties. We highlight that unbounded multipliers fail in this setting, as the bootstrapped eigenvalues asymptotically follow a Fréchet or Gumbel distribution. Second, for spiked sample covariance matrices, while both bounded and unbounded multipliers can recover the asymptotic normality of the largest eigenvalues, they may introduce additional bias, particularly when the spikes are not strong. To mitigate this, we apply a modified multiplier bootstrap multiple times to correct the bias. Finally, leveraging our modified multiplier bootstrap procedures, we propose a novel and straightforward distribution-based test for selecting common factors in the factor model. Numerical simulations validate the accuracy and robustness of our proposed methods, demonstrating superior performance compared to existing approaches in the literature. Technically, we establish the asymptotic distributions of the largest eigenvalues of bootstrapped sample covariance matrices for various classes of multipliers in both spiked and non-spiked models, which may be of independent interest.

1. Introduction

The bootstrap method [30] is a powerful and widely used tool in multivariate statistics and machine learning. By resampling a single dataset to generate numerous simulated samples, it facilitates inference even when little is known about the properties of the data-generating distribution. This approach is particularly appealing when theoretical derivations based on asymptotic analysis are complex or rely on restrictive assumptions. On a related note, the covariance matrix plays a central role in virtually every aspect of multivariate data analysis. Over the past few decades, technological advancements have spurred growing interest in developing methodologies and tools to address the challenges posed by high-dimensionality and complexity [73]. In particular, extreme eigenvalues of sample covariance matrices are critical in principal component analysis (PCA) [1, 48]. However, current theoretical findings

MSC2020 subject classifications: Primary 60B20, 62G10; secondary 15B52, 62H15.
Keywords and phrases: Sample covariance matrix, Edge eigenvalues, Multiplier bootstrap, Spike detection.

on these extreme eigenvalues often depend on unknown parameters of the population covariance matrix and are typically intricate [8, 73], rendering conventional methods impractical in many cases.

An intuitive approach to overcome the above challenges is to apply bootstrap methodology to covariance matrices in the high-dimensional regime. Then, a natural question arises:

Could the asymptotic distribution of extreme sample eigenvalues be effectively approximated using the multiplier bootstrap method? If not, can the multiplier bootstrap be adapted to maintain its efficiency in high-dimensional settings for PCA?

Addressing this question is highly non-trivial, as the extreme sample eigenvalues of large covariance matrices exhibit complex limiting distributions that depend on the structure of the population covariance matrix. For example, non-spiked covariance matrices often follow the Tracy-Widom distribution [9, 26, 28, 39, 32, 47, 51, 55, 65], whereas spiked covariance matrices tend to follow some Gaussian distribution [7, 8, 15, 48, 63]. In all these distributions, the asymptotic results often involve unknown and complex quantities, making it challenging to apply these results in practice. Moreover, as highlighted in [34, 49], directly applying standard bootstrap methods in high-dimensional regimes can sometimes yield erratic results for statistical inference involving extreme eigenvalues. This issue becomes particularly pronounced in scenarios where population spikes are weak or entirely absent; see Section 1.1 below for further discussion of these challenges.

Motivated by these difficulties, this paper provides a comprehensive analysis of the effects of multiplier bootstrap procedures on the asymptotic behavior of the top eigenvalues of sample covariance matrices in high-dimensional settings. It also proposes practical and effective methods for specific statistical tasks using multiplier bootstrap techniques. We demonstrate that preserving information from the original sample covariance matrix through multiplier bootstrap procedures poses significant challenges. These challenges often necessitate careful selection of multipliers, a large number of bootstrap replications, and, in some cases, additional bias corrections to improve statistical estimates after bootstrapping.

Before going to the details, we first generate our bootstrap procedure for high-dimensional sample covariance matrices as follow. Consider a sequence of data $\mathbf{s}_i \sim \mathbf{s} \in \mathbb{R}^p, 1 \leqslant i \leqslant n$, which are i.i.d. observations of a random vector $\mathbf{s}$ such that

$$\mathbf{s} = \Sigma^{1/2}\mathbf{x} \in \mathbb{R}^p, \tag{1.1}$$

where $\Sigma \in \mathbb{R}^{p\times p}$ is deterministic representing the covariance structures in the dataset and $\mathbf{x} \in \mathbb{R}^p$ is a random vector containing i.i.d. centered random variables with variance n^{-1}. For high-dimensionality, we mean that p and n are comparably large. Now, given a sequence of data $\mathbf{s}_i = \Sigma^{1/2}\mathbf{x}_i, 1 \leq i \leq n$, we resample $\mathbf{s}_i$'s via a sequence of random multipliers $\xi_i \sim \xi \in \mathbb{R}$ which are independent with $\mathbf{x}$. The resampled data can be described as

$$\mathbf{y}_i = \xi_i \Sigma^{1/2}\mathbf{x}_i \in \mathbb{R}^p, \quad 1 \leq i \leq n. \tag{1.2}$$

We may write the resampled data matrix as $Y = \Sigma^{1/2}XD$, where $X = (\mathbf{x}_i)$ and D is a diagonal matrix containing $\{\xi_i\}$. Then the bootstrapped sample covariance matrix can be constructed as follows

$$Q := YY^* \equiv \Sigma^{1/2}XD^2X^*\Sigma^{1/2}. \tag{1.3}$$

Technically, understanding the asymptotic behavior of the largest eigenvalues of Q is crucial for analyzing the performance of the multiplier bootstrap for the edge eigenvalues of the sample covariance matrix. In what follows, we first provide a summary of some related results in Section 1.1. Then we offer an overview of our contributions in Section 1.2.

1.1. Summary of some existing related results

In this subsection, we summarize results related to the bootstrap methodology. While the bootstrap is extensively studied in the literature [3, 12, 18, 30, 31, 66], our focus is specifically on aspects relevant to sample covariance matrices. The use of the bootstrap for studying sample covariance matrices dates back to [11, 20, 29] and has since evolved into a powerful tool in multivariate analysis [1, 18, 61]. Key findings from these studies show that nonparametric bootstrap methods can effectively approximate the eigenvalue distribution of sample covariance matrices in low-dimensional settings, where the sample size approaches infinity while the data dimension remains fixed or grows slowly.

More recently, the research has focused on evaluating whether the bootstrap methods can reliably capture the asymptotic properties of sample covariance matrices in high-dimensional settings. We summarize some closely related literature as follows. For the global behavior of the spectrum of sample covariance matrices, [57] investigated the problem of bootstrapping linear spectral statistics for datasets with the structure (1.2). Subsequently, [69] extended this study to the bootstrap of linear spectral statistics in the high-dimensional elliptical model. Moreover, [58] explored the efficiency of bootstrapping the operator norm under various population decay profiles, while [75] developed a universal bootstrap statistic based on the covariance operator norm for testing covariance matrices. Additionally, [19] proposed a nonparametric sampling-with-replacement bootstrap for eigenvalue statistics of high-dimensional sample covariance matrices.

Regarding individual eigenvalues, much less attention has been given to high-dimensional settings, except for a few cases under certain structural assumptions. These assumptions generally ensure that the individual eigenvalues of sample covariance matrices exhibit Gaussian behavior. For instance, [41] studied the multiplier bootstrap for the largest eigenvalue in a moderately diverging dimension setting, assuming $p = \mathrm{o}(n^{1/9})$, while [72] investigated the standard bootstrap for the largest eigenvalue, assuming that the eigenvalues of Σ decay exponentially. Similar assumptions and results for eigenvectors were established in [60]. More recently, [74] examined the standard bootstrap under a factor model, assuming strong factor strength. However, it remains unclear and challenging to determine whether the multiplier bootstrap can perform effectively in high-dimensional settings without relying on such strong structural assumptions, as questioned in [34, 49].

Finally, we note that (1.3) is often referred to as a *separable sample covariance matrix* in the context of random matrix theory (RMT). In the literature, such models have been studied primarily under scenarios where both Σ and D are bounded and deterministic; see, for instance, [17, 28, 33, 64, 71, 76]. However, our focus on the random matrix model in (1.3) differs from the aforementioned studies, as we treat D^2 as random multipliers, whose range may also be unbounded. Existing results addressing the case of random D are limited [33, 76], and these are confined to analyzing the limiting spectral distribution under specific conditions. In this regard, our findings contribute to the RMT literature by providing the limiting distributions of the edge eigenvalues of a novel class of separable sample covariance matrices with random structures, which may be of independent interest.

1.2. An overview of our results and contributions

In this section, we provide an informal overview of our results and highlight the main contributions and novelties of our paper. At a high level, our findings and proposed algorithms demonstrate that the multiplier bootstrap can effectively analyze the asymptotics of the largest few eigenvalues of sample covariance matrices, whether spiked or not, provided that the multipliers are appropriately chosen and the bootstrap procedures are applied multiple times. We elaborate on this in more detail below.

In Section 3, we examine the feasibility of the multiplier bootstrap for non-spiked sample covariance matrices in high dimensions. First, Theorem 3.1 reveals that the multiplier bootstrap fails to replicate the asymptotic Tracy-Widom distribution for the largest eigenvalues of non-spiked sample covariance matrices when the multipliers are unbounded. This failure arises because, in such a setting, the largest eigenvalues of the bootstrapped sample covariance matrices instead follow either a Fréchet or Gumbel distribution. Second, in contrast, Theorem 3.3 shows that for various types of bounded multipliers, the largest eigenvalues of the bootstrapped sample covariance matrices can follow Tracy-Widom, Gaussian, or Weibull distributions. This makes it possible to recover the asymptotic distribution of the largest eigenvalues of the sample covariance matrices with appropriately chosen bounded multipliers, provided that the parameters of the Tracy-Widom law are accurately estimated. Third, to achieve this, we propose a new modified procedure, Algorithm 1, which generates a large number of bootstrapped sample covariance matrices using carefully designed bounded multipliers. Theoretically, Corollary 3.6 establishes that our modified multiplier bootstrap procedure can effectively capture the Tracy-Widom law for non-spiked sample covariance matrices—a task previously considered challenging in [34].

In Section 4, we examine the effectiveness of the multiplier bootstrap for spiked sample covariance matrices. As established in Theorem 4.1, the largest eigenvalues of the bootstrapped sample covariance matrices under the spiked model are asymptotically Gaussian, regardless of whether the multipliers are bounded, under mild assumptions. While this suggests that the multiplier bootstrap may recover the asymptotic normality of the largest eigenvalues of the spiked covariance matrix, it also introduces a biased mean and a more intricate variance structure, as highlighted in Theorems 4.1 and 4.3, particularly when the spikes are not sufficiently strong. Direct theoretical estimation of the asymptotic mean and variance is highly challenging, requiring stronger assumptions on the spikes and additional technical efforts [34, 72]. To address these issues, we propose a novel algorithm, Algorithm 2, which utilizes multiple bootstrapped sample covariance matrices with carefully chosen multipliers whose effect remains dominated by the actual spikes (cf. (4.1)). Corollary 4.5 demonstrates that our refined multiplier bootstrap procedure effectively captures the asymptotics of the leading spiked eigenvalues of the sample covariance matrices under weaker assumptions for the spikes.

Finally, in Section 5, we explore the application of the multiplier bootstrap to PCA for selecting common factors, building on our established results. Technically, this problem reduces to detecting spikes in a spiked covariance matrix model. Specifically, we propose a novel distribution-based test using the multiplier bootstrap procedure to test and estimate the number of spikes. Our new algorithm, Algorithm 3, is simple and leverages the results and methods developed in Sections 3 and 4. The key idea is to run Algorithm 2 multiple times with suitably chosen unbounded multipliers and then perform a normality test on the resulting statistics under a well-established hypothesis (cf. (5.2)). This approach is motivated by our theoretical findings: under the null hypothesis (i.e., some spikes exist), Corollary 4.5 guarantees that the output statistics from Algorithm 2 are asymptotically Gaussian. However, under the alternative hypothesis, Theorem 3.1 implies that the statistics follow either a Fréchet or Gumbel distribution. Consequently, a simple normality test enables us to detect and consistently estimate the number of spikes (cf. (5.3)).

We note that, technically, this problem reduces to studying the largest eigenvalues of the bootstrapped sample covariance matrices (1.3) under various assumptions on the multipliers, for both spiked and non-spiked models. Our analysis reveals that the asymptotic distribution of these eigenvalues can take various forms, including the three extreme value distributions for sequences of i.i.d. random variables—Gumbel, Fréchet, and Weibull [10]—as well as the Tracy-Widom (TW) law, Gaussian, or a mixture of TW and Gaussian distributions.

The specific distribution depends on D^2 (i.e., the multipliers), Σ, the presence of spikes, and the aspect ratio p/n. These theoretical findings are of independent interest, offering insights into how bootstrap mechanisms influence the spectral limits of covariance matrices. Consequently, they provide guidance for designing appropriate multiplier mechanisms to effectively bootstrap the largest eigenvalues in both spiked and non-spiked models for various statistical applications.

The rest of this article is organized as follows. In Section 2, we give the details of our model and some basic assumptions. In Section 3, we present the main results of multiplier bootstrap for the non-spiked covariance matrix model. In Section 4, we study the multiplier bootstrap for the spiked covariance matrix model. In Section 5, we consider application of multiplier bootstrap methodologies in common factor selection. Numerical simulations are also provided to show the usefulness of our algorithm. Some preliminary and proof strategies are summarized in Section 6. Technique proof and details are deferred to our supplementary material [25].

Conventions. Let $\mathbb{C}_+$ be the complex upper half plane. We denote $C>0$ as a generic constant whose value may change from line to line. For two sequences of deterministic positive values $\{a_n\}$ and $\{b_n\}$, we write $a_n=\mathrm{O}(b_n)$ if $a_n \leqslant Cb_n$ for some positive constant $C>0$. In addition, if both $a_n=\mathrm{O}(b_n)$ and $b_n=\mathrm{O}(a_n)$, we write $a_n \asymp b_n$. Moreover, we write $a_n=\mathrm{o}(b_n)$ if $a_n \leqslant c_n b_n$ for some positive sequence $c_n \downarrow 0$. Moreover, for a sequence of random variables $\{x_n\}$ and positive real values $\{a_n\}$, we use $x_n=\mathrm{O}_{\mathbb{P}}(a_n)$ to state that x_n/a_n is stochastically bounded. Similarly, we use $x_n=\mathrm{o}_{\mathbb{P}}(a_n)$ to say that x_n/a_n converges to zero in probability. For a sequence of positive random variables $\{y_n\}$, we use $y_{(k)}, 1 \leqslant k \leqslant n$, for its order statistics with $y_{(1)} \geqslant y_{(2)} \geqslant \cdots \geqslant y_{(n)} > 0$.

2. The model and basic assumptions

In this section, we introduce our model and some assumptions. As discussed around (1.3), we consider bootstrapped data matrix of the following form

$$Y=\Sigma^{1/2} X D, \tag{2.1}$$

where Σ is a $p\times p$ deterministic positive definite matrix, D is an $n \times n$ diagonal random weights matrix containing i.i.d. multipliers, and X is a $p \times n$ random matrix independent of D whose entries satisfying the following assumption.

ASSUMPTION 2.1. Throughout the paper, we assume that the entries of $X=(x_{ij})$ are centered i.i.d. random variables satisfying that for $1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n$,

$$\mathbb{E} x_{ij}=0,\ \mathbb{E} x_{ij}^2=\frac{1}{n}. \tag{2.2}$$

Moreover, we assume that for all $k \in \mathbb{N}$, there exists some constant $C_k>0$ so that $\mathbb{E}|\sqrt{n} x_{ij}|^k \leqslant C_k$.

For the multipliers, we impose the following mild assumptions.

ASSUMPTION 2.2. Let $D^2=\operatorname{diag}\left\{\xi_1^2, \cdots, \xi_n^2\right\}$. Moreover, for its entries, we assume $\xi_i^2 \sim \xi^2, 1 \leqslant i \leqslant n$, are i.i.d. generated from a nonnegative and non-degenerated random variable ξ^2 satisfying the following assumptions.

(i) **Unbounded support case.** We assume that ξ^2 has an unbounded support and satisfies either of the following two conditions:
(a). ξ^2 is a regularly varying random variable [67] that

$$\mathbb{P}(\xi^2 > x) = \frac{L(x)}{x^{\alpha}}, \quad x \to \infty, \tag{2.3}$$

for some $\alpha \in [2, +\infty)$, where $L(x)$ is a slowly varying function in the sense that for all $t > 0$, $\lim_{x\to\infty} L(tx)/L(x) = 1$.
(b). ξ^2 has an exponential decay tail in the sense that for some constant $\beta > 0$ and any fixed constant $t > 0$

$$\mathbb{E}e^{t\xi^{2\beta}} < \infty. \tag{2.4}$$

(ii) **Bounded support case**. We assume that ξ^2 has a bounded support on $(0, l]$ for fixed some constant $l > 0$. Moreover, for some constant $d > -1$, we assume that

$$\mathbb{P}(l - \xi^2 \leqslant x) \asymp x^{d+1}. \tag{2.5}$$

Finally, let $F(x)$ be the cumulative distribution function (CDF) of ξ^2, we assume that

$$0 < \mathfrak{b} := \lim_{x \uparrow l} \frac{1 - F(x)}{(l-x)^{d+1}} < \infty. \tag{2.6}$$

REMARK 2.3. Several remarks are in order. First, for the unbounded random multipliers, (2.3) indicates that the tails of ξ^2 decay polynomially. Many commonly used distributions are included in this category. To name but a few, Pareto distribution, F distribution and student-t distribution. Moreover, according to extreme value theory (see Lemma S.1.16 of our supplement), when (2.3) is satisfied, $\xi^2_{(1)}$ follows Fréchet distribution asymptotically. Second, for the unbounded setting, (2.4) implies that the tails of ξ^2 decay exponentially. In fact, by elementary calculations [42], it is not hard to see that when (2.4) holds, it is necessarily that the CDF of ξ^2 admits

$$\mathbb{P}(\xi^2 > x) = \exp(-\mathsf{g}(x)), \tag{2.7}$$

for some positive decreasing function $\mathsf{g}(x) > 0$. Furthermore, if

$$\mathsf{g} \in C^{\infty}([0, \infty)), \quad \lim_{x \uparrow \infty} (1/\mathsf{g}'(x))' = 0, \tag{2.8}$$

we see from Lemma S.1.16 of our supplement that $\xi^2_{(1)}$ follows Gumbel distribution asymptotically. In fact, many commonly used distributions, for instance, Chi-squared distribution, exponential distribution and Gamma distribution, satisfy these conditions.

Third, for the bounded random multipliers, (2.5) indicates that ξ^2 has a possible polynomial decay behavior near the edge. Under the assumption of (2.6), we see from Lemma S.1.16 of our supplement that $\xi^2_{(1)}$ obeys Weibull distribution asymptotically. The conditions allow for many distributions like (shifted) Beta distribution, uniform distribution and U-quadratic distribution. In summary, we emphasize that our assumptions in Assumption 2.2 are general and mild and cover many commonly used multipliers.

The following assumption introduces some mild conditions on the aspect ratio p/n and the population covariance matrix Σ.

ASSUMPTION 2.4. We assume the following conditions hold true for some small constant $0 < \tau < 1$.

(i) **On dimensionality**. Throughout the paper, we consider the high dimensional regime that

$$\tau \leqslant \phi := \frac{p}{n} \leqslant \tau^{-1}. \tag{2.9}$$

(ii) **On** Σ. For the population covariance matrix Σ, we assume that it admits the following spectral decomposition

$$\Sigma = \sum_{j=1}^{p} \sigma_j \mathbf{v}_j \mathbf{v}_j^*, \tag{2.10}$$

where

$$\tau \leqslant \sigma_p \leqslant \sigma_{p-1} \leqslant \cdots \leqslant \sigma_2 \leqslant \sigma_1 \leqslant \tau^{-1}, \tag{2.11}$$

are the eigenvalues and $\{\mathbf{v}_j\}$ are the associated eigenvectors.

We remark that (2.9) is commonly used in random matrix theory and high dimensional statistics literature for quantifying the high dimensionality. (2.11) states the eigenvalues of the population covariance matrix are bounded from above and below. On the one hand, when ξ^2 has unbounded support as in Case (i) of Assumption 2.2, (2.11) is the only assumption imposed on Σ. On the other hand, when ξ^2 has bounded support as in Case (ii) of Assumption 2.2, we will provide an additional mild assumption, Assumption S.1.1 of our supplement, to exclude potential spikes.

In the statistical literature, motivated by real applications, one often adds some spikes to Σ which results in the famous spiked covariance matrix model [21, 47]. To construct such a model, one can introduce a perturbed version of Σ, denoted as $\widetilde{\Sigma}$ whose spectral decomposition follows

$$\widetilde{\Sigma} = \sum_{j=1}^{p} \widetilde{\sigma}_j \mathbf{v}_j \mathbf{v}_j^*, \tag{2.12}$$

where for some fixed constant $r > 0$, $\widetilde{\sigma}_1 \geqslant \widetilde{\sigma}_2 \geqslant \cdots \geqslant \widetilde{\sigma}_r > \widetilde{\sigma}_{r+1}$ are r values representing the larger spikes while the rest $\widetilde{\sigma}_j = \sigma_j, j \geqslant r+1$'s are relatively small and bounded. For simplicity and definiteness, we assume that for some constant $\tau > 0$

$$\frac{\widetilde{\sigma}_i}{\widetilde{\sigma}_{i+1}} \geqslant 1 + \tau, \ 1 \leqslant i \leqslant r. \tag{2.13}$$

Based on $\widetilde{\Sigma}$, the counterpart of bootstrapped data matrix (2.1) can be written as

$$\widetilde{Y} = \widetilde{\Sigma}^{1/2} X D.$$

Consequently, the bootstrapped sample covariance matrix (i.e., the counterpart of (1.3)) can be written as

$$\widetilde{Q} := \widetilde{Y}\widetilde{Y}^* \equiv \widetilde{\Sigma}^{1/2} X D^2 X^* \widetilde{\Sigma}^{1/2}. \tag{2.14}$$

As explained earlier, the understanding of the multiplier boostrap on the edge eigenvalues of sample covariance matrices boils down to the study of the first few largest eigenvalues of the $p \times p$ bootstrapped sample covariance matrices Q in (1.3) or $\widetilde{Q}$ in (2.14). To clarify the notations used for the various matrices, we summarize them in Table 1.

Model	Quantity	Sample version	Bootstrapped version
Non-spiked	Matrix	$S := \Sigma^{1/2} X X^* \Sigma^{1/2}$	$Q := \Sigma^{1/2} X D^2 X^* \Sigma^{1/2}$
	Eigenvalues	$\{\widehat{\lambda}_i\}$	$\{\lambda_i\}$
Spiked	Matrix	$\widetilde{S} := \widetilde{\Sigma}^{1/2} X X^* \widetilde{\Sigma}^{1/2}$	$\widetilde{Q} := \widetilde{\Sigma}^{1/2} X D^2 X^* \widetilde{\Sigma}^{1/2}$
	Eigenvalues	$\{\widehat{\mu}_i\}$	$\{\mu_i\}$

TABLE 1
Summary of some important notations.

3. Multiplier bootstrap meets the non-spiked covariance matrix model

In this section, we present the first part of the main results by evaluating the effectiveness of the bootstrap method for different classes of multipliers. Specifically, we establish the asymptotic distributions of the largest eigenvalues of the bootstrapped sample covariance matrix (1.3) when the population covariance matrix Σ lacks significant large spikes. In this scenario, the extreme eigenvalues of the sample covariance matrix S in Table 1 follow the Tracy-Widom distribution.

In Section 3.1, we demonstrate that when the multipliers are unbounded, the bootstrap method becomes invalid. Subsequently, in Section 3.2, we show that the Tracy-Widom law can be recovered by carefully selecting appropriate multipliers. To achieve this, we propose a new algorithm, Algorithm 1, specifically designed to address this challenge in Section 3.3.

3.1. The bad: unbounded multipliers are invalid

We first provide the results for the extreme eigenvalues when the multiplier ξ^2 has unbounded support in the sense that (i) of Assumption 2.2 holds. Denote

$$\bar{\sigma}_1 = \frac{1}{p}\sum_{i=1}^{p} \sigma_i, \quad \bar{\sigma}_2 = \frac{1}{p}\sum_{i=1}^{p} \sigma_i^2.$$

Recall $F(x)$ is the cumulative distribution function (CDF) of $\xi_i^2, 1 \leqslant i \leqslant n$. Denote

$$b_n := \inf\left\{ x : 1 - F(x) \leqslant \frac{1}{n} \right\}. \tag{3.1}$$

Recall from Table 1 that λ_1 is the largest eigenvalue of Q.

THEOREM 3.1 (Unbounded multipliers). *Suppose Assumptions 2.1, 2.4 and (i) of Assumption 2.2 hold. Then we have that when n is sufficiently large*

$$\frac{\lambda_1}{\xi_{(1)}^2} = \varphi + \mathrm{o}_{\mathbb{P}}(1). \tag{3.2}$$

where for ϕ in (2.9), $\varphi := \phi\bar{\sigma}_1$. Consequently, when (2.3) holds, $\varphi^{-1}\lambda_1$ follows the Fréchet distribution asymptotically in the sense that for $x \geqslant 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\lambda_1}{\varphi b_n} \leqslant x \right) = \exp\left(-x^{-\alpha}\right). \tag{3.3}$$

Moreover, when (2.4) and (2.8) hold, $\varphi^{-1}\lambda_1$ follows the Gumbel distribution asymptotically in the sense that for $x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(\mathsf{g}'(b_n)\left[\varphi^{-1}\lambda_1 - (b_n + c_0)\right] \leqslant x \right) = \exp\left(-e^{-x}\right), \tag{3.4}$$

where we recall g *is defined in (2.7) and* $c_0 := 1 + \varphi^{-1} \times \mathbb{E}\xi^2 \times \bar{\sigma}_2/\bar{\sigma}_1$.

REMARK 3.2. Two remarks are in order. First, Theorem 3.1 states that when ξ^2 has unbounded support, λ_1 will be divergent. Moreover, after being properly centered and scaled, λ_1 will have a similar behavior to $\xi^2_{(1)}$. Especially, when ξ^2 has a polynomial decay tail as in (2.3), we can obtain the Fréchet limit and when ξ^2 has an exponential decay tail as in (2.4), we can get the Gumbel limit. Second, Theorem 3.1 also indicates that if we select unbounded multipliers, the typical Tracy-Widom (TW) limit of largest eigenvalues from sample covariance matrices has no chance to be reproduced. Third, the above results can be generalized to the joint distribution of k largest eigenvalues for any fixed k. That is, for all $s_i \in \mathbb{R}, 1 \leqslant i \leqslant k$, (3.3) can be generalized to

$$\lim_{n\to\infty} \mathbb{P}\left(\left(\frac{\lambda_i}{\varphi b_n} \leqslant s_i\right)_{1\leqslant i\leqslant k}\right) = \lim_{n\to\infty} \mathbb{P}\left(\left(\frac{\xi^2_{(i)}}{b_n} \leqslant s_i\right)_{1\leqslant i\leqslant k}\right),$$

and (3.4) can be generalized to

$$\lim_{n\to\infty} \mathbb{P}\left(\mathsf{g}'(b_n)\left(\varphi^{-1}\lambda_i - (b_n + c_0) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right) = \lim_{n\to\infty} \mathbb{P}\left(\mathsf{g}'(b_n)\left(\xi^2_{(i)} - (b_n + c_0) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right).$$

Since the joint distribution of the order statistics of $\{\xi_i^2\}$ can be computed explicitly [16], the above formulas give a complete description of the finite-dimensional correlation functions of the extremal eigenvalues. Finally, we mention that the Fréchet distribution and Gumbel distribution also appear in the literature in heavy-tailed sample covariance matrices, see [4, 43, 44, 45] for example.

3.2. The good: recovering TW law is possible with bounded multipliers

Next, we state the results when the multiplier ξ^2 has bounded support in the sense that (ii) of Assumption 2.2 holds. Recall $F(x)$ and l from (2.6), and using the definitions of $m_{1n,c}$ and L_+ from (6.3) and (6.4) below, we denote

$$\mathsf{s}_1 := \int_0^l \frac{l^2 s^2}{(l-s)^2} \mathrm{d}F(s), \quad \mathsf{s}_2 := \int_0^l \frac{ls}{l-s} \mathrm{d}F(s),$$
$$\mathsf{s}_3 := \frac{1}{p}\sum_{i=1}^p \frac{\sigma_i^2 \mathsf{s}_1}{(L_+ - \sigma_i \mathsf{s}_2)^2}, \quad \mathsf{s}_4 := \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{(-L_+ + \sigma_i \mathsf{s}_2)^2}, \tag{3.5}$$

and

$$\mathsf{v} := \int \left(\frac{s}{1 + s m_{1n,c}(L_+)}\right)^2 \mathrm{d}F(s) - \left(\int \frac{s}{1 + s m_{1n,c}(L_+)} \mathrm{d}F(s)\right)^2.$$

Recall the exponent d in (2.5).

THEOREM 3.3 (Bounded multipliers). *Suppose Assumptions 2.1, 2.4, S.1.1 and (ii) of Assumption 2.2 hold. Then we have that when* n *is sufficiently large,*

(1). *When* $d > 1$ *and* $\phi^{-1} > \mathsf{s}_3$, *we have that* L_+ *satisfies*

$$1 = \frac{1}{n}\sum_{i=1}^p \frac{-l\sigma_i}{-L_+ + \sigma_i \mathsf{s}_2}. \tag{3.6}$$

Moreover, we have that

$$n^{\frac{1}{d+1}} \left| \left(\frac{\lambda_1 - L_+}{\mathsf{s}_4^{-1}(1-\phi \mathsf{s}_3)} \right) - \left(\xi_{(1)}^2 - l \right) \right| = \mathrm{o}_{\mathbb{P}}(1). \tag{3.7}$$

Consequently, we have that $\lambda_1 - L_+$ follows Weibull distribution with parameter $d+1$ asymptotically in the sense that for $x \leqslant 0$

$$\lim_{n \to \infty} \mathbb{P} \left(\frac{(\mathfrak{b} n)^{d+1}}{\mathsf{s}_4^{-1}(1-\phi \mathsf{s}_3)} (\lambda_1 - L_+) \leqslant x \right) = \exp\left(-|x|^{d+1} \right), \tag{3.8}$$

where $\mathfrak{b}$ is defined in (2.6).

(2). *When $d > 1$ and $\phi^{-1} < \mathsf{s}_3$, we have that λ_1 is asymptotically Gaussian in the sense that*

$$\lim_{n \to \infty} \mathbb{P} \left(\sqrt{n \mathsf{v}^{-1}} (\lambda_1 - L_+) \leqslant x \right) = \Phi(x), \tag{3.9}$$

where $\Phi(x)$ is the CDF of a real standard Gaussian random variable.

(3). *When $-1 < d \leqslant 1$, we have that*

$$\lambda_1 - L_+ = \nu_1 + \nu_2 + \mathrm{O}_{\mathbb{P}}(n^{-1}),$$

where for some γ defined in (S.23) of our supplement, $n^{2/3}\gamma\nu_1$ follows the type-1 Tracy-Widom law asymptotically and $\sqrt{n}\nu_2/\mathsf{v}^{1/2}$ follows standard Gaussian distribution asymptotically. More specifically, if $\mathsf{v} = \mathrm{o}(n^{-1/3})$,

$$\lim_{n \to \infty} \mathbb{P} \left(n^{2/3} \gamma (\lambda_1 - L_+) \leqslant x \right) = \mathrm{T}(x),$$

where $\mathrm{T}(x)$ is the CDF of the type-1 Tracy-Widom distribution. Moreover, if $\mathsf{v} \gg n^{-1/3}$, (3.9) holds.

REMARK 3.4. Theorem 3.3 shows that, when ξ^2 has bounded support as in (2.5), λ_1 will be bounded and can have several phase transitions depending on the exponent d, aspect ratio ϕ and the threshold s_3 which encodes the information of Σ and the distribution of ξ^2.

First, in the setting when $d > 1$, on the one hand, when $\phi^{-1} > \mathsf{s}_3$, after being properly centered and scaled, λ_1 will have similar asymptotics as $\xi_{(1)}^2$ and Weibull limit will be obtained. On the other hand when $\phi^{-1} < \mathsf{s}_3$, λ_1 will be influenced by all $\{\xi_i^2\}$ and hence asymptotically Gaussian. For the critical case $\phi^{-1} = \mathsf{s}_3$, we believe there will be a phase transition connecting Gaussian and Weibull. Since this is out of the scope of the paper which focuses on statistical applications, we will pursue this direction in the future works.

Second, when $-1 < d \leqslant 1$, the limiting ESD of Q (cf. ρ in Theorem 6.2) will have a square root decay behavior. In this setting, λ_1 will be influenced by two components, the TW part ν_1 and the Gaussian part ν_2. The TW part is due to the square root behavior and the Gaussian is due to the fact that λ_1 will be potentially influenced by all $\{\xi_i^2\}$; see Section S.4.2 of our supplement for more details. We mention that the variance of the Gaussian part can potentially decay and ν_1 and ν_2 are in generally dependent. Moreover, by appropriately choosing the multipliers to ensure the variance of the Gaussian component diminishes, the fluctuation of $\lambda_1 - L_+$ is predominantly governed by the Tracy-Widom distribution. This approach demonstrates that the Tracy-Widom distribution can be recovered using the bootstrap method with carefully selected bounded multipliers, as detailed in the following subsection.

Finally, as discussed in Remark 3.2, we can generalize the results of Theorem 3.3 to the joint distribution of k largest eigenvalues for any fixed k. We omit the details.

3.3. The modification: adjusted multiplier bootstrap for the TW law

In this subsection, we explore an application of Theorem 3.3 to recover the TW distribution, which was previously considered a pessimistic task in [34] using the multiplier bootstrap. As discussed in Remark 3.4, the theoretical insights behind (3) of Theorem 3.3 suggest that it is, in principle, feasible to recover the TW law by appropriately selecting multipliers ξ^2 with sufficiently strong concentration properties, such that the quantity v in Theorem 3.3 decays significantly faster than $\mathrm{O}(n^{-1/3})$. Inspired by [46], a promising candidate is $\xi^2 = T^{-1}\sum_{t=1}^{T}\zeta_t^2$, where T is some sufficiently large number and the ζ_t^2's are i.i.d. random variables satisfying

$$\mathbb{E}(\zeta_t^2) = 1, \quad \mathbb{E}(\zeta_t^4) < \infty.$$

For simplicity, we can take the ζ_t's as i.i.d. Gaussian random variables. In this case, it is straightforward to verify that $\mathbb{E}(\xi^2) = 1$ and $\mathrm{Var}(\xi^2) = 2T^{-1}$. More importantly, ξ^2 can be shown to exhibit strong convergence around its mean. As a result, for large value of T, we can have that $\mathsf{v} = \mathrm{o}(n^{-1/3})$, and the largest eigenvalue of the bootstrapped matrix will follow the TW limit as described in (3) of Theorem 3.3.

Below, we propose a novel algorithm to implement the above idea for recovering the TW distribution by bootstrapping the data matrix $\Sigma^{1/2}X$. The key idea is to employ appropriate multipliers and generate multiple bootstrapped sample covariance matrices.

Algorithm 1 Adjusted multiplier bootstrap for TW law

Inputs: The data matrix $\Sigma^{1/2}X$, the size $T = \lfloor n/2 \rfloor$ and number of resampling $B = \lfloor n^{5/3} \rfloor$.
Step One: For each $j = 1, \ldots, n$, generate T i.i.d. copies of the Gaussian random variable ζ, say $\{\zeta_{jt}\}_{1 \le t \le T}$. Set $\xi_j^2 = T^{-1}\sum_{t=1}^{T}\zeta_{jt}^2$ and construct $D^2 = \mathrm{diag}(\xi_1^2, \ldots, \xi_n^2)$.
Step Two: Repeat the above procedures B times to obtained a sequence of multiplier matrices $\{D_b^2\}_{1 \le b \le B}$. Construct the bootstrapped sample covariance matrices from data matrix $\Sigma^{1/2}X$ as $Q_b = \Sigma^{1/2}XD_b^2X^*\Sigma^{1/2}, 1 \leqslant b \leqslant B$.
Step Three: Compute the largest eigenvalue of each $Q_b, 1 \le b \le B$, denoted as $\lambda_{b,1}, 1 \le b \le B$. Compute the estimator $\widetilde{L}_+ = B^{-1}\sum_{b=1}^{B}\lambda_{b,1}$.
Output: The empirical distribution (conditional on the data) based on $n^{2/3}\{\lambda_{b,1} - \widetilde{L}_+\}_{1 \le b \le B}$, denoted as $F_{\mathsf{TW}}(x)$.

REMARK 3.5. Two remarks are in order. First, in Algorithm 1, the value of T must be chosen sufficiently large to ensure that the TW fluctuations dominate the limiting behavior of each $\lambda_{b,1}$. Second, accurately estimating the unknown parameter L_+ requires performing the bootstrap procedure a sufficiently large number of times—significantly exceeding $n^{4/3}$—to reduce the estimation error below the fluctuation level of the TW distribution, which is $n^{-2/3}$.

The theoretical analysis of Algorithm 1 is summarized in the following corollary, the proof of which will be provided in Section S.4 of our supplement. This corollary demonstrates that our proposed Algorithm 1 is capable of recovering the asymptotic distribution of $\widehat{\lambda}_1$ (recall from Table 1 that $\widehat{\lambda}_1$ denotes the largest eigenvalue of the sample covariance matrix).

COROLLARY 3.6. *Recall the notations in Table 1. Suppose the assumptions of Theorem 3.3 hold. Then conditional on the data $\Sigma^{1/2}X$, we have that for all $x \in \mathbb{R}$*

$$\lim_{n\to\infty}\mathbb{P}(n^{2/3}(\widehat{\lambda}_1 - E_+) \le x) = \lim_{n\to\infty}F_{\mathsf{TW}}(x), \tag{3.10}$$

where E_+ is the rightmost edge of the limiting spectral distribution of S which is usually unknown in practice.

As established in the literature (e.g., [9, 26, 28, 32, 47, 51, 55, 65]), the left-hand side of (3.10) follows a (scaled) TW law. Combined with Theorem 3.3 and Remark 3.4, the right-hand side is also shown to asymptotically follow a (scaled) TW law. More importantly, Corollary 3.6 makes the application of the TW law practical, as it eliminates the need to estimate the unknown quantity E_+. Finally, we note that Algorithm 1 and Corollary 3.6 can be extended to the first k largest eigenvalues for any fixed k, as discussed in Remark 3.4. Furthermore, they can be applied to test the structure of Σ. For a detailed discussion, we refer readers to Section S.4.4 of our supplement.

4. Multiplier bootstrap meets the spiked covariance matrix model

In this section, we present the second part of our results. Specifically, we derive the asymptotic distributions of the largest eigenvalues of the bootstrapped sample covariance matrix (1.3) when the population covariance matrix Σ exhibits a prominent spiked structure. Under this setting, the extreme eigenvalues of the sample covariance matrices $\widetilde{S}$ in Table 1 follow a Gaussian distribution with complex variance structures.

In Section 4.1, we demonstrate that the bootstrap method is generally valid, apart from a bias component, when the population spikes exceed a certain threshold. Building on this, in Section 4.2, we propose a novel bias correction procedure to enhance the efficiency and accuracy of the multiplier bootstrap

4.1. The good: multiplier bootstrap can be useful for the spikes

In this section, we demonstrate that the multiplier bootstrap can be a valuable tool for analyzing the asymptotics of large eigenvalues in sample covariance matrices with spiked structures. Specifically, when suitably chosen multipliers are applied, the leading eigenvalues of the bootstrapped sample covariance matrix corresponding to the population spikes retain Gaussian distributions, making it possible to study the asymptotics of the largest eigenvalues of the sample covariance matrices. To formalize this, we introduce the following threshold for various multipliers according to Assumption 2.2

$$\mathsf{T} := \begin{cases} n^{1/\alpha}\log n, & \text{if (2.3) holds;} \\ \log^{1/\beta} n, & \text{if (2.4) holds;} \\ l, & \text{if (2.5) holds.} \end{cases} \tag{4.1}$$

In fact, T serves as a reference point for identifying suitable multipliers, determined by the spike strength of $\widetilde{\Sigma}$ as in (2.12). For each $1 \le i \le r$, we denote a deterministic quantity θ_i which is the unique solution of the equation,

$$\frac{\theta_i}{\widetilde{\sigma}_i} = \left(1 - \frac{1}{n\theta_i}\sum_{j=r+1}^{p} \frac{\sigma_j}{1-\widetilde{\sigma}_i^{-1}\sigma_j}\right)^{-1},$$

with restriction $\theta_i \in [\widetilde{\sigma}_i, 2\widetilde{\sigma}_i]$. Furthermore, we denote

$$\mathsf{M}_i := \mathbb{E}\xi^2\left[\left(\frac{1}{n}\sum_{k=r+1}^{p}\frac{\sigma_k}{\widetilde{\sigma}_i}\times\mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^2\right)+\left((\frac{1}{n}\sum_{k=r+1}^{p}\frac{\sigma_k}{\widetilde{\sigma}_i})^2\times\mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^3\right)\right], \tag{4.2}$$

and

$$\mathsf{V}_i := \frac{1}{n}\sum_{j=1}^{n}\left[\sum_{k=r+1}^{p} v_{ki}^4(\mathfrak{m}_4-3)+3(\sum_{k=r+1}^{p} v_{ki}^2)^2\right]-1, \tag{4.3}$$

where $\mathbf{v}_i=(v_{1i},\ldots,v_{pi})^*$ is the eigenvector of $\widetilde{\Sigma}$ and $\mathfrak{m}_4=\mathbb{E}(\sqrt{n}x_{11})^4$.

Recall from Table 1 that $\{\mu_i\}$ are the eigenvalues of $\widetilde{Q}$.

THEOREM 4.1. *Suppose Assumptions 2.1, 2.2 and 2.4 hold. For the spikes in (2.12) and (2.13), we assume*

$$\widetilde{\sigma}_r \gg \mathsf{T}. \tag{4.4}$$

Then conditional on the data matrix $\widetilde{\Sigma}^{1/2}X$, we have for $1\leqslant i\leqslant r$ and $x\in\mathbb{R}$,

$$\lim_{n\to\infty}\mathbb{P}\left(\sqrt{\frac{n}{\mathsf{V}_i}}\left(\frac{\mu_i}{\theta_i}-\mathsf{M}_i\right)\leqslant x\right)=\Phi(x), \tag{4.5}$$

where M_i and V_i are defined in (4.2) and (4.3), respectively, and $\Phi(x)$ is the CDF of the standard Gaussian random variable.

REMARK 4.2. Two remarks are in order. First, the condition (4.4) can be verified in practice. Recall from Table 1 that $\{\widehat{\mu}_i\}$ are the eigenvalues of the spiked sample covariance matrices $\widetilde{S}$. Specifically, according to [15], for $1\leqslant i\leqslant r$, it holds that $\widehat{\mu}_i/\widetilde{\sigma}_i=1+\mathrm{o}_{\mathbb{P}}(1)$. Consequently, $\widehat{\mu}_r$ can serve as a proxy for $\widetilde{\sigma}_r$, offering valuable guidance for selecting suitable multipliers based on T.

Second, as noted in Table 1, the non-spiked eigenvalues of $\widetilde{Q}$ closely follow those of the non-spiked matrix Q. Specifically, for any fixed integer k, we can show that

$$|\mu_{r+i}-\lambda_i|=\mathrm{O}_{\mathbb{P}}\left(n^{-1/2+2\epsilon}d_1\right),\quad 1\leqslant i\leqslant k, \tag{4.6}$$

where d_1 is a slowly divergent constant defined in (S.4) of our supplement. By combining (4.6) with Remark 3.2, we conclude that μ_{r+i} $(1\leqslant i\leqslant k)$ follows either a Fréchet or Gumbel distribution, depending on the tail behavior of the multiplier. In contrast, Theorem 4.1 shows that the spiked eigenvalues are always Gaussian. This distinction highlights the differing distributions of the spiked and non-spiked eigenvalues in the bootstrapped sample covariance matrix, providing a theoretical foundation for detecting spikes. This aspect will be explored further in Section 5.

It is important to note that the results in Theorem 4.1 cannot be directly applied in practice since all θ_i, M_i and V_i are unknown. In what follows, we show that θ_i can be replaced by the eigenvalues of the sample covariance matrix (recall Table 1).

THEOREM 4.3. *Under assumptions of Theorem 4.1, we have for $1\leq i\leq r$,*

$$\lim_{n\to\infty}\mathbb{P}\Big(\sqrt{\frac{n}{\mathsf{V}_i+1}}\Big(\frac{\mu_i}{\widehat{\mu}_i}-\mathsf{M}_i\Big)\leq x\Big)=\Phi(x), \tag{4.7}$$

where M_i and V_i are same as in Theorem 4.1.

REMARK 4.4. The CLT result in Theorem 4.3 highlights that while the multiplier bootstrap methodology is possibly useful for spiked sample covariance matrices, it generally introduces bias parts, as reflected in M_i and V_i. This sheds light on why the bootstrap technique may exhibit instability in approximating the distribution of the top eigenvalues across almost all types of multipliers, as observed in [49].

In pratice, both M_i and V_i are often hard to be estimated. In the literature, several technical assumptions are commonly employed to ease this difficulty, such as assuming the population covariance is isotropic [5, 6, 37, 68] or requiring the spike strength to exceed the threshold of $n^{1/2}$ [72, 74]. However, leveraging the intrinsic nature of the bootstrap methodology, one can approximate M_i and V_i directly through asymptotic normality by performing the multiplier bootstrap procedure multiple times. The details of this idea will be presented in the next subsection.

4.2. The modification: adjusting multiplier bootstrap via bias correction

In this subsection, we leverage Theorem 4.3 to develop an algorithm for estimating the asymptotic mean and variance of $\mu_i/\widehat{\mu}_i$ for $1 \leqslant i \leqslant r$ so that the modified multiplier bootstrap can be applied. The idea is similar to Algorithm 1 by running the multiplier bootstrap procedure multiple times using some well-chosen multipliers. The algorithm below encapsulates this approach.

Algorithm 2 Adjusted multiplier bootstrap for spiked eigenvalues

Inputs: The data matrix $\widetilde{\Sigma}^{1/2}X$, the first r sample eigenvalues $\{\widehat{\mu}_i\}_{1\leq i\leq r}$, the number of resampling $B = \lfloor n^{3/2} \rfloor$.
Step One: Choose ξ^2 from Assumption 2.2 satisfying $\mathsf{T} \ll \widehat{\mu}_r$. Generate B multiplier matrices $\{D_b^2\}_{1\leq b\leq B}$ and construct the bootstrapped covariance matrices
$\widetilde{Q}_b = \widetilde{\Sigma}^{1/2} X D_b^2 X^* \widetilde{\Sigma}^{1/2}, 1 \leqslant b \leqslant B$.
Step Two: Compute the top r largest eigenvalue of each $\widetilde{Q}_b, 1 \leq b \leq B$, denoted as $\mu_{b,i}, 1 \leq b \leq B, 1 \leq i \leq r$. For each $1 \leq i \leq r$, computer the estimators
$\widehat{\mathsf{M}}_i = B^{-1} \sum_{b=1}^{B} \mu_{b,i}/\widehat{\mu}_i$ and $\widehat{\mathsf{V}}_i = (B-1)^{-1} \sum_{b=1}^{B} (\mu_{b,i}/\widehat{\mu}_i - \widehat{\mathsf{M}}_i)^2$.
Output: The empirical distributions (conditional on the data) based on
$n^{1/2}\{(\mu_{b,i}/\widehat{\mu}_i - \widehat{\mathsf{M}}_i)/\sqrt{\widehat{\mathsf{V}}_i}\}_{1\leq b\leq B}$, for each $1 \leq i \leq r$, denoted as $F_{\mathsf{G}}^{(i)}(x)$.

Similar to the discussions in Remark 3.5, the number of times of multiplier bootstrap procedures should be sufficiently large, for instance, $B \gg n$, to minimize the estimation error of $\widehat{\mathsf{M}}_i$ and $\widehat{\mathsf{V}}_i$ relative to the Gaussian scaling of $n^{1/2}$. For simplicity, we choose $B = \lfloor n^{3/2} \rfloor$. The theoretical properties of Algorithm 2 can be summarized as follows.

COROLLARY 4.5. *Under assumptions in Theorem 4.3, we have for $1 \leq i \leq r$ and all $x \in \mathbb{R}$*

$$\lim_{n\to\infty} F_{\mathsf{G}}^{(i)}(x) = \Phi(x),$$

where $\Phi(x)$ is the CDF of the standard Gaussian random variable.

Corollary 4.5 provides an application for spike detection. Specifically, if the population covariance matrix $\widetilde{\Sigma}$ contains r spikes, the first r empirical distributions $F_{\mathsf{G}}^{(i)}$, obtained through Algorithm 2, will approximately follow a standard Gaussian distribution, even when the support of the multipliers is unbounded. In contrast, for indices $i \geqslant r+1$, the distributions $F_{\mathsf{G}}^{(i)}$ will conform to (rescaled) extreme value distributions, as directly implied by (4.6) and Theorem 3.1. We explore this application and demonstrate its effectiveness in Section 5.

5. Statistical applications in common factors selection

In this section, we explore the application of the modified multiplier bootstrap methodology for selecting common factors in the factor model, drawing on the algorithms and results presented in Sections 3 and 4.

Suppose we have obtained the data matrix $Z=(\mathbf{z}_i)$, where $\mathbf{z}_i, 1 \leqslant i \leqslant n$, are i.i.d. sampled from the factor model

$$\mathbf{z} = L\mathbf{f} + \mathbf{e} \in \mathbb{R}^p, \tag{5.1}$$

where $\mathbf{f}$ is an $r \times 1$ low rank (unobserved) factor, L is a $p \times r$ low rank loading matrix and $\mathbf{e}$ is the $p \times 1$ idiosyncratic error which is independent of $\mathbf{f}$. For the purpose of identifiability, following [5, 6, 37, 68], we assume that $\operatorname{Cov}(\mathbf{f}, \mathbf{f}) = I_r$. Therefore, the covariance structure of $\mathbf{z}$ can be written as $\widetilde{\Sigma} := LL^* + \operatorname{Cov}(\mathbf{e}, \mathbf{e})$. Thanks to its low-rank structure, in high dimension, $\widetilde{\Sigma}$ is often assumed to follow a spiked model, as described in (2.12) and (2.13), [15, 37, 38].

In factor model applications, a key question is determining the number of common factors needed to explain the economic variables. This can be framed as identifying the number of spikes, r, in the data matrix Z. More explicitly, we are interested in testing the value of r via the hypothesis that

$$\mathbf{H}_0 : r \geqslant r_0 \ \text{ vs } \ \mathbf{H}_a : r < r_0, \tag{5.2}$$

where r_0 is some pre-given integer representing our belief of the value of r. Based on it, we can further propose the sequential testing estimator for r as

$$\widehat{r} := \sup \left\{ r_0 \geqslant 0 : \ \mathbf{H}_0 \text{ is accepted} \right\}. \tag{5.3}$$

In the literature, many methods based on the eigenvalues of the sample covariance matrix ZZ^* have been proposed for the hypothesis testing problem (5.2) in terms of factor models under our setting, for example, see [2, 5, 15, 37, 62]. However, most of the existing results in the literature rely on specific structural assumptions about Z or employ complex techniques to predict its limiting spectral behavior. In contrast, we propose a novel and straightforward distribution-biased test for (5.2), leveraging our modified multiplier bootstrap methods outlined in Algorithms 1 and 2. In the following, we write $W := ZD^2Z^*$ as the bootstrapped matrix of ZZ^*, and we abusively denote their eigenvalues as $\{\widehat{\mu}_i\}$ (from ZZ^*) and $\{\mu_i\}$ (from W), respectively.

Intuitively, Corollary 4.5 suggests that under the null hypothesis, the statistics used in Algorithm 2 are asymptotically Gaussian when conditioned on the data, provided that appropriately chosen multipliers (both bounded and unbounded) are used. Furthermore, Theorem 3.1 indicates that when the multipliers have unbounded supports, the statistics from Algorithm 2 follow an extreme value distribution rather than a Gaussian distribution. Consequently, by employing well-chosen unbounded multipliers and running Algorithm 2, we can perform the Kolmogorov-Smirnov test for normality to evaluate (5.2). The algorithm is presented below.

Algorithm 3 Multiplier Bootstrapped Distribution Testing for (5.2)

Inputs: Pre-chosen integer r_0, data matrix Z, type I error α and the critical value $\mathsf{D}_{n,\alpha}$ from one-sample Kolmogorov-Smirnov table.
Step One: Compute the sample eigenvalues of ZZ^*, say $\{\widehat{\mu}_i\}_{1\leq i\leq n}$. Choose the distribution of unbounded multiplier ξ^2 by $\mathsf{T} \ll \widehat{\mu}_{r_0}$.
Step Two: Input data matrix Z, our chosen unbounded multiplier into Algorithm 2 and run for the largest r_0-th sample eigenvalues. Obtain the output $F_{\mathsf{G}}^{(r_0)}(x)$.
Step Three: Compute the Kolmogorov-Smirnov statistics $D_{n,r_0} = \sup_x |F_{\mathsf{G}}^{(r_0)}(x) - \Phi(x)|$ and conduct the Kolmogorov-Smirnov test.
Output: Reject $\mathbf{H}_0$ in (5.2) if $D_{n,r_0} > \mathsf{D}_{n,\alpha}$.

Generally speaking, under $\mathbf{H}_0$, $\{F_{\mathsf{G}}^{(i)}(x)\}_{1\leq i\leq r_0}$ are asymptotically standard Gaussian distribution from Corollary 4.5, while $F_{\mathsf{G}}^{(r_0)}(x)$ will follow either a Fréchet or Gumbel distribution under $\mathbf{H}_a$. We summarize the theoretical properties of Algorithm 3 as follows.

COROLLARY 5.1. *Suppose the assumptions of Corollary 4.5 hold, given some significant level α,*

$$\lim_{n\to\infty} \mathbb{P}(D_{n,r_0} \leq \mathsf{D}_{n,\alpha}) = 1-\alpha, \quad \textit{under } \mathbf{H}_0;$$

while

$$\lim_{n\to\infty} \mathbb{P}(D_{n,r_0} > \mathsf{D}_{n,\alpha}) = 1, \quad \textit{under } \mathbf{H}_a.$$

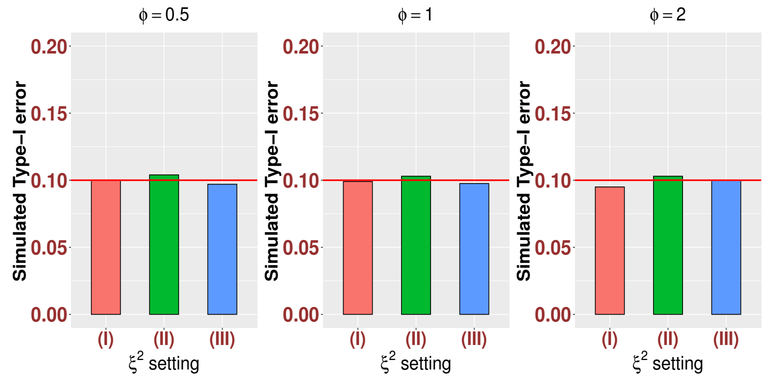
In what follows, we conduct Monte-Carlo simulations to demonstrate the accuracy, power and robustness of our proposed Algorithm 3 under the factor model setup (5.1). For simplicity, considering the setups in [37, 74], in the simulations, for the data matrix $Z \in \mathbb{R}^{p\times n}$, we assume that

$$Z = \delta L'\mathbf{F} + \mathbf{E},$$

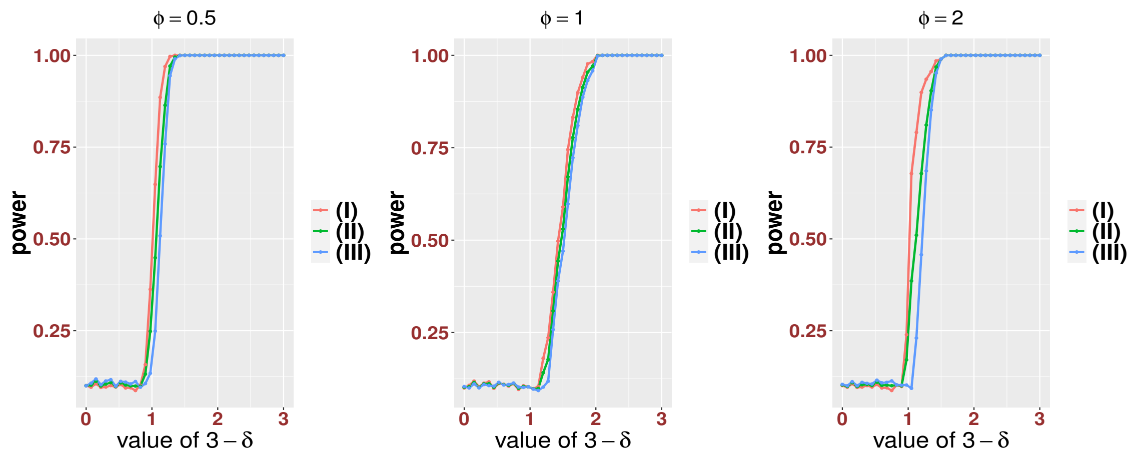
where $L' \in \mathbb{R}^{p\times 3}$ is the loading matrix whose rows are independent Gaussian random vectors in $\mathbb{R}^3$ with covariance matrix $\mathrm{diag}\{1.3, 0.8, 0.5\}$, $\mathbf{F} \in \mathbb{R}^{3\times n}$ is the factor score matrix independent of L' with i.i.d. standard Gaussian entries and $\mathbf{E} \in \mathbb{R}^{p\times n}$ is a standard Gaussian matrix independent of the factor loading and score matrices. Here $\delta \geqslant 0$ is the factor strength. Under this setup, the null of (5.2) can be characterized as $\mathbf{H}_0 : r = 3$ which reduces to checking whether δ is large enough. The alternative of (5.2) can be expressed as $\mathbf{H}_a : r = 0$ which reduces to checking whether $\delta = 0$.

First, we study our proposed statistics. We check the accuracy under $\alpha = 0.1$ under the null that $r = 3$ with $\delta = 3$. Moreover, we also examine the power of the statistics for the alternative when $r = 0$ which implies $\delta = 0$. We can conclude from Figure 1 that our Algorithm 3 is reasonably accurate and powerful for various choices of multipliers ξ^2 under different settings of ϕ.

Second, we compare the performance of our approaches with some existing ones. Again, since most of the existing literature focus on the estimation of the number of the spikes instead of inferring, we compare the performance of our inference based estimator $\widehat{r}$ in (5.3) with a few existing ones for estimating the number of factors in the context of factor model. For definiteness, we compare our estimators with the ones proposed in [2, 5, 15, 37, 62]. In Figure 2, we compare the accuracy of our estimators when $r = 3$ using correct detection ratio (CDR). We can find that our estimators can outperform some of the existing ones especially when the spikes (i.e., factor strengths) are not that large.



(a) Simulated type I error.



(b) Simulated power.

FIG 1. *Simulated type I error rates and power under the nominal level 0.1 for our proposed Algorithm 3. Here we consider three different settings of multiplier* ξ^2 *: (I). Gamma distribution with parameters 15 and 15, (II).* $\exp(1)$ *distribution, and (III).* χ_1^2 *distribution. We take* $n = 400$ *and report the results based on 2,000 Monte-Carlo simulations. The randomness in* Z *are i.i.d. Gaussian with mean zero and variance* n^{-1}.

6. Some background and strategies for the proof

In this section, we present a concise overview of the proof strategies for the main results in Sections 3 and 4. To precisely analyze the bootstrap effect on the fluctuation of the largest eigenvalue of the sample covariance matrix, we leverage and further develop techniques from random matrix theory to quantify the interaction between the randomness introduced by the multipliers and the observed data samples. At a high level, guided by concepts from free probability theory [59], this interaction can be interpreted as the convolution between the multiplier matrix D and the data matrix X. To clarify the technical details, we begin by introducing several fundamental notations and concepts from the random matrix theory literature in Section 6.1, followed by an explanation of our proof strategies in Section 6.2.

6.1. Asymptotic laws

In this section, we introduce some results on the limiting asymptotic laws of the eigenvalues of the bootstrapped sample covariance matrices Q in (1.3) with its $n \times n$ companion $\mathcal{Q} :=$

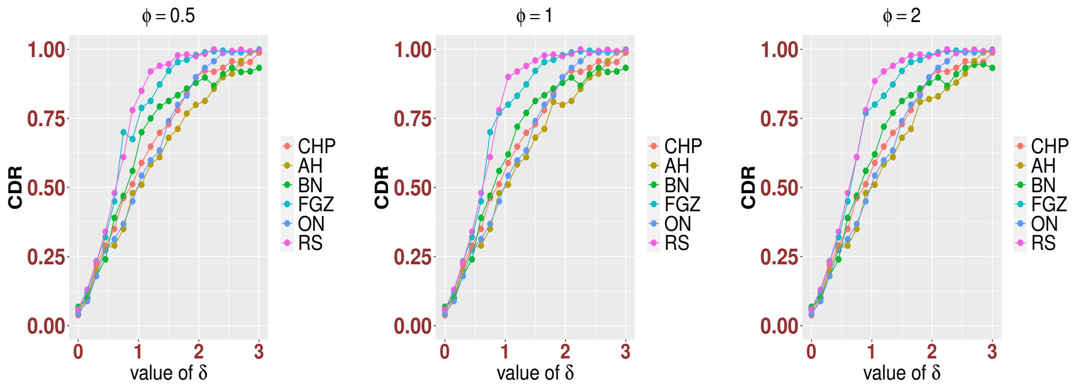


FIG 2. *Comparison of estimation. In the above figures, "CHP" refers to the method from [15], "AH" refers to the method from [2], "BN" refers to the method from [5], "FGZ" refers to the method from [37], "ON" refers to the method from [62], and "RS" refers to our proposed method in Algorithm 3 where the entries of* D^2 *are i.i.d.* $\exp(1)$ *random variables. Here* $n = 400$ *and the entries of* X *are i.i.d. Gaussian with mean zero and variance* n^{-1}*. The CDR is reported using 2,000 simulations.*

$DX^*\Sigma XD$. Recall that the empirical spectral distributions (ESD) of Q and $\mathcal{Q}$ are denoted as

$$\mu_Q := \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i(Q)}, \quad \mu_{\mathcal{Q}} := \frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j(\mathcal{Q})}.$$

It is well-known that the ESDs can be best described via its the Stieltjes transforms as follows

$$m_Q := \int \frac{1}{x-z}\mu_Q, \quad m_{\mathcal{Q}} := \int \frac{1}{x-z}\mu_{\mathcal{Q}},\ z \in \mathbb{C}_+. \tag{6.1}$$

Since Q and $\mathcal{Q}$ share the same non-trivial eigenvalues, it suffices to study μ_Q and m_Q. The limit of μ_Q can be described by a system of equations [17, 27, 33, 46, 64, 76]. To avoid repetitions, we summarize these equations in the following definition.

DEFINITION 6.1 (Systems of consistent equations). *For $z \in \mathbb{C}_+$, we define the triplets $(m_{1n}, m_{2n}, m_n) \in \mathbb{C}_+^3$, via the following systems of equations.*

$$m_{1n}(z) = \frac{1}{n}\sum_{i=1}^{p}\frac{\sigma_i}{-z(1+\sigma_i m_{2n}(z))}, \quad m_{2n}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i^2}{-z(1+\xi_i^2 m_{1n}(z))}, \tag{6.2}$$

$$m_n(z) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{-z(1+\sigma_i m_{2n}(z))}.$$

For sufficiently large n, we find that μ_Q has a nonrandom deterministic equivalent and can be uniquely characterized by the above consistent equations. This is summarized by the following theorem whose proof can be obtained by following lines of the arguments of [33, Theorem 2] and [64, Theorem 1] verbatim.

THEOREM 6.2 (Asymptotic laws). *Suppose Assumptions 2.1, 2.2 and 2.4 hold. Then conditional on some event $\Omega \equiv \Omega_n$ that $\mathbb{P}(\Omega) = 1 - \mathrm{o}(1)$, for any $z \in \mathbb{C}_+$, when n is sufficiently large, there exists a unique solution $(m_{1n}(z), m_{2n}(z), m_n(z)) \in \mathbb{C}_+^3$ to the systems of equations in (6.2). Moreover, $m_n(z)$ is the Stieltjes transform of some probability density function $\rho \equiv \rho_n$ defined on $\mathbb{R}$ which can be obtained using the inversion formula.*

REMARK 6.3. Several remarks on Theorem 6.2 are in order. First, the probability event Ω can be constructed explicitly as in Definition S.1.10 and Lemma S.1.12 in our supplement [25]. Second, we prove an unconditional counterpart for Theorem 6.2 by integrating out the randomness of multiplier ξ^2. Recall $F(x)$ is the CDF of ξ^2. We can define the counterpart for (6.2) as follows

$$m_{1n,c}(z) = \frac{1}{n}\sum_{i=1}^{p}\frac{\sigma_i}{-z(1+\sigma_i m_{2n,c}(z))}, \quad m_{2n,c}(z) = \int_0^l \frac{s}{-z(1+s m_{1n,c}(z))}\mathrm{d}F(s),$$

$$m_{n,c}(z) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{-z(1+\sigma_i m_{2n,c}(z))}. \tag{6.3}$$

In this setting, $(m_{1n,c}, m_{2n,c}, m_{n,c})$ is always deterministic. Especially, when ξ^2 has bounded support as in Case (ii) of Assumption 2.2, we can actually obtain stronger results as in [64] that the support of the associated probability density function $\widetilde{\rho}$ is bounded and denoted as

$$\operatorname{supp}(\widetilde{\rho}) = [L_-, L_+]. \tag{6.4}$$

6.2. Sketch of the proof strategies

In this section, we outline the proof strategies for Theorems 3.1, 3.3, and 4.1, focusing on the largest eigenvalue. We start with the non-spiked model (Section 3). In this setting, the arguments differ for bounded and unbounded multipliers D. On the one hand, when D has unbounded support (Theorem 3.1), the eigenvalues will be divergent. In this setting, we utilize a perturbation argument. However, as in this case, the associated ρ in Theorem 6.2 may also be unbounded, the perturbation approach developed in [14, 27, 50] cannot be applied directly. Instead, we modify the perturbation arguments by isolating $\mathbf{y}_i$ corresponding to the largest multiplier $\xi^2_{(1)}$ from the bootstrapped matrix Y as in (1.3). More explicitly, for the proof of Theorem 3.1, with $m_{1n}(z)$ in (6.2), the key is to introduce a real auxiliary quantity $\vartheta_1 > 0$ to be the largest solution of

$$1 + (\xi^2_{(1)} + d_1)m_{1n}(\vartheta_1) = 0, \tag{6.5}$$

where $d_1 = \mathrm{o}_{\mathbb{P}}(\xi^2_{(1)})$ (cf. (S.4) of our supplement) is introduced for some technical reasons. First, as will be seen in our proofs (cf. (S.3) and (S.8) of our supplement), (6.5) provides a natural way to connect ϑ_1 and $\xi^2_{(1)}$ in the sense that $\vartheta_1/\xi^2_{(1)} = \varphi + \mathrm{o}_{\mathbb{P}}(1)$. Secondly, it establishes a connection between ϑ_1 and λ_1 through a modified perturbation argument based on [14, 21, 27]. Specifically, λ_1 can be uniquely characterized by the equation $M(\lambda_1) = 0$ (cf. (S.5) of our supplement), where $M(\cdot)$ (cf. (S.4) of our supplement) is a random quantity isolating the column in (2.1) associated with $\xi^2_{(1)}$. Using our newly established local laws (cf. Theorem S.1.8 of our supplement), we can further demonstrate that $1 + (\xi^2_{(1)} + d_1)m_{1n}(\lambda_1) \approx 0$. Subsequently, a detailed continuity and stability analysis shows that $\lambda_1/\vartheta_1 = 1 + \mathrm{o}_{\mathbb{P}}(1)$, thereby completing the proof of Theorem 3.1.

On the other hand, when the multiplier D has bounded support (Theorem 3.3), our arguments are non-perturbative and extend the approach introduced in [52, 56]. This generalization is non-trivial and requires a dedicated analysis of the relationship between multipliers and the largest eigenvalues. Specifically, it necessitates a sophisticated understanding of the systems of equations, such as those in (6.2), on local scales. A key input is the distinct local behaviors of the asymptotic law (cf. $\widetilde{\rho}(x)$ in (6.4)) near the edge under varying settings (cf. Lemma S.1.5 of our supplement). More precisely, we find that $\widetilde{\rho}(x) \sim \sqrt{L_+ - x}$ under the setup of (2) and (3) of Theorem 3.3, and $\widetilde{\rho}(x) \sim (L_+ - x)^d$ under (1) of Theorem 3.3.

In the actual proof for Theorem 3.3, we decompose λ_1 into two components: $\lambda_1 - \widehat{L}_+$ and $\widehat{L}_+ - L_+$, where the quantity $\widehat{L}_+$ is defined as the edge of the conditional density function ρ in Theorem 6.2 by fixing a realization of the multipliers $\{\xi_i^2\}$ on Ω. For cases (2) and (3) of Theorem 3.3, where the square root behavior emerges, $n^{2/3}(\lambda_1 - \widehat{L}_+)$ asymptotically follows the Tracy–Widom (TW) law with constant-order variance. Regarding the unconditional distribution, the fluctuation of $\widehat{L}_+$, due to the i.i.d. assumption of D^2, is asymptotically Gaussian by the Central Limit Theorem (CLT), with a variance that may decay to zero. Consequently, the overall distribution can be expressed as a sum of the TW law and a Gaussian component (potentially with vanishing variance). For case (1) of Theorem 3.3, the key observation is that $\widehat{L}_+$ can be represented as the solution of

$$m_{1n}(\widehat{L}_+) = -l^{-1}. \tag{6.6}$$

Moreover, for $z \in \mathbb{C}_+$ in a small neighborhood of $\widehat{L}_+$, $m_{1n}(z)$ can be expanded linearly as $m_{1n}(z) - m_{1n}(\widehat{L}_+) = \beta(z - \widehat{L}_+) + \mathrm{o}(n^{-1/(d+1)})$ for some constant β (cf. Lemma S.1.5 of our supplement). Then, it allows us to locate λ_1 neighboring around $\widehat{L}_+$ as

$$\operatorname{Re} m_{1n}(\lambda_1 + \mathrm{i}\eta_0) \approx -\xi_{(1)}^{-2}, \quad \eta_0 = n^{-1/2-\epsilon_d}. \tag{6.7}$$

Connecting (6.7) with $l - \xi_{(1)}^2$ and (6.6), we can conclude (3.7) for $\widehat{L}_+$. For the unconditional result with L_+, it follows directly from Lemma S.1.5 of our supplement that $L_+ = \widehat{L}_+ + \mathrm{O}_{\mathbb{P}}(n^{-1/2+\delta})$. Since $d > 1$, we can conclude that λ_1 is only influenced by $\xi_{(1)}^2$ and asymptotically Weibull.

We then discuss the proof of the spiked model (Section 4). Given the spiked structure in (2.12), our proof primarily relies on a perturbative argument tailored for divergent spikes, as developed in [15]. However, the inclusion of multipliers introduces additional complexity, requiring a generalization of the above perturbative techniques. Due to similarity, our discussion focuses on Theorem 4.1. A key element of our proof is the introduction of a random term, ζ_1 (cf. (S.8) of our supplement), which captures the randomness of μ_1/θ_1 solely through the randomness of $\{\xi_i\}$. The asymptotic Gaussianity arises from the asymptotic normality of ζ_1, which can be established via a standard central limit theorem (CLT) argument. To quantify the bias term M_1 in (4.2), we carefully analyze $\zeta_1 - \theta_1/\widetilde{\sigma}_1$. This requires a more refined analysis of the local scales of the systems defined in Definition 6.1, leveraging our newly established local laws in Theorems S.1.8 and S.1.9 of our supplement.

Acknowledgments

XCD wants thank Zhigang Bao, Miles Lopes and Fan Yang for many helpful discussions. XCD is supported in part by funds from NSF DMS 2113489 and DMS 2306439.

REFERENCES

[1] ABDI, H. and WILLIAMS, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* **2** 433–459.

[2] AHN, S. C. and HORENSTEIN, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81** 1203–1227.

[3] ALEMAYEHU, D. (1988). Bootstrapping the latent roots of certain random matrices. *Communications in Statistics-Simulation and Computation* **17** 857–869.

[4] AUFFINGER, A., BEN AROUS, G. and PÉCHÉ, S. (2009). Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annales de l'IHP Probabilités et statistiques* **45** 589–610.

[5] BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221.

[6] BAI, J., NG, S. et al. (2008). Large dimensional factor analysis. *Foundations and Trends® in Econometrics* **3** 89–163.

[7] BAI, Z. and YAO, J.-F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'IHP Probabilités et statistiques* **44** 447–474.

[8] BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics* **50** 1144–1169.

[9] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics* **43** 382–421.

[10] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. L. (2004). *Statistics of extremes: theory and applications* **558**. John Wiley & Sons.

[11] BERAN, R. and SRIVASTAVA, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics* **13** 95–115.

[12] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9** 1196–1217.

[13] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electronic Journal of Probability* **19** 1–53.

[14] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probability theory and related fields* **164** 459–552.

[15] CAI, T. T., HAN, X. and PAN, G. (2020). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. *The Annals of Statistics* **48** 1255 – 1280.

[16] COLES, S. (2001). *An introduction to statistical modeling of extreme values*. *Springer Series in Statistics*. Springer.

[17] COUILLET, R. and HACHEM, W. (2014). Analysis of the limiting spectral measure of large random matrices of the separable covariance type. *Random Matrices: Theory and Applications* **3** 1450016.

[18] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap methods and their application*. Cambridge university press.

[19] DETTE, H. and ROHDE, A. (2024). Nonparametric bootstrap of high-dimensional sample covariance matrices. *arXiv preprint arXiv:2406.16849*.

[20] DIACONIS, P. and EFRON, B. (1983). Computer-intensive methods in statistics. *Scientific American* **248** 116–131.

[21] DING, X. (2021). Spiked sample covariance matrices with possibly multiple bulk components. *Random Matrices: Theory and Applications* **10** 2150014.

[22] DING, X. and JI, H. C. (2023). Local laws for multiplication of random matrices. *The Annals of Applied Probability* **33** 2981–3009.

[23] DING, X. and JI, H. C. (2023). Spiked multiplicative random matrices and principal components. *arXiv preprint arXiv:2302.13502*.

[24] DING, X. and XIE, J. (2023). Tracy-Widom distribution for the edge eigenvalues of elliptical model. *arXiv preprint arXiv 2304.07893*.

[25] DING, X., XIE, J., YU, L. and ZHOU, W. (2025). Supplement to "Multiplier bootstrap meets high-dimensional PCA: the good, the bad and the modification".

[26] DING, X. and YANG, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *The Annals of Applied Probability* **28** 1679–1738.

[27] DING, X. and YANG, F. (2021). Spiked separable covariance matrices and principal components. *The Annals of Statistics* **49** 1113–1138.

[28] DING, X. and YANG, F. (2022). Tracy-Widom Distribution for Heterogeneous Gram Matrices With Applications in Signal Detection. *IEEE Transactions on Information Theory* **68** 6682-6715.

[29] EATON, M. L. and TYLER, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *The Annals of Statistics* 260–271.

[30] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7** 1–26.

[31] EFRON, B. (1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics* **9** 139–158.

[32] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability* **35** 663–714.

[33] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability* **19** 2362–2405.

[34] EL KAROUI, N. and PURDOM, E. (2019). The non-parametric bootstrap and spectral analysis in moderate and high-dimension. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2115–2124.

[35] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability* **41** 2279–2375.

[36] ERDŐS, L. and YAU, H.-T. (2017). *A dynamical approach to random matrix theory*. American Mathematical Soc.

[37] FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *Journal of the American Statistical Association* **117** 852–861.

[38] FAN, J., WANG, K., ZHONG, Y. and ZHU, Z. (2021). Robust high dimensional factor models with applications to statistical machine learning. *Statistical science* **36** 303.

[39] FAN, Z. and JOHNSTONE, I. M. (2022). Tracy-Widom at each edge of real covariance and MANOVA estimators. *The Annals of Applied Probability* **32** 2967.

[40] FANG, K.-T. and ANDERSON, T. W. (1990). *Statistical inference in elliptically contoured and related distributions*. Allerton Press.

[41] HAN, F., XU, S. and ZHOU, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli* **24** 1787 – 1833.

[42] HANSEN, A. (2020). The three extreme value distributions: An introductory review. *Frontiers in Physics* **8** 604053.

[43] HEINY, J. and MIKOSCH, T. (2017). Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Processes and their Applications* **127** 2179–2207.

[44] HEINY, J. and MIKOSCH, T. (2021). Large sample autocovariance matrices of linear processes with heavy tails. *Stochastic Processes and their Applications* **141** 344–375.

[45] HEINY, J., MIKOSCH, T. and YSLAS, J. (2021). Point process convergence for the off-diagonal entries of sample covariance matrices. *The Annals of Applied Probability* **31** 538 – 560.

[46] HU, J., LI, W. and ZHOU, W. (2019). Central limit theorem for mutual information of large MIMO systems with elliptically correlated channels. *IEEE Transactions on Information Theory* **65** 7168–7180.

[47] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics* **29** 295–327.

[48] JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proceedings of the IEEE* **106** 1277–1292.

[49] KAROUI, N. E. and PURDOM, E. (2016). The bootstrap, covariance matrices and PCA in moderate and high-dimensions. *arXiv preprint arXiv:1608.00948*.

[50] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Communications on Pure and Applied Mathematics* **66** 1663–1749.

[51] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields* **169** 257–352.

[52] KWAK, J., LEE, J. O. and PARK, J. (2021). Extremal eigenvalues of sample covariance matrices with general population. *Bernoulli* **27** 2740–2765.

[53] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of statistics* 1302–1338.

[54] LEE, J. O. and SCHNELLI, K. (2013). Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *Journal of Mathematical Physics* **54** 103504.

[55] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *The Annals of Applied Probability* **26** 3786–3839.

[56] LEE, J. O. and SCHNELLI, K. (2016). Extremal eigenvalues and eigenvectors of deformed Wigner matrices. *Probability Theory and Related Fields* **164** 165–241.

[57] LOPES, M. E., BLANDINO, A. and AUE, A. (2019). Bootstrapping spectral statistics in high dimensions. *Biometrika* **106** 781–801.

[58] LOPES, M. E., ERICHSON, N. B. and MAHONEY, M. W. (2023). Bootstrapping the operator norm in high dimensions: Error estimation for covariance matrices and sketching. *Bernoulli* **29** 428–450.

[59] MINGO, J. A. and SPEICHER, R. (2017). *Free probability and random matrices* **35**. Springer.

[60] NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2019). Bootstrap confidence sets for spectral projectors of sample covariance. *Probability Theory and Related Fields* **174** 1091–1132.

[61] OLIVE, D. J., OLIVE, D. J. and CHERNYK (2017). *Robust multivariate analysis*. Springer.

[62] ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* **92** 1004–1016.

[63] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* 1617–1642.

[64] PAUL, D. and SILVERSTEIN, J. W. (2009). No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis* **100** 37–57.

[65] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *The Annals of Applied Probability* **24** 935–1001.

[66] POLITIS, D. N., ROMANO, J. P., WOLF, M., POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling in the IID Case*. Springer.
[67] RESNICK, S. I. (2008). *Extreme values, regular variation, and point processes* **4**. Springer Science & Business Media.
[68] STOCK, J. H. and WATSON, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, **2** 415–525.
[69] WANG, S. and LOPES, M. E. (2023). A bootstrap method for spectral statistics in high-dimensional elliptical models. *Electronic Journal of Statistics* **17** 1848–1892.
[70] WEN, J., XIE, J., YU, L. and ZHOU, W. (2022). Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions. *Bernoulli* **28** 2941–2967.
[71] YANG, F. (2019). Edge universality of separable covariance matrices. *Electronic Journal of Probability* **24** 1–57.
[72] YAO, J. and LOPES, M. E. (2023). Rates of bootstrap approximation for eigenvalues in high-dimensional PCA. *Statistica Sinica* **33** 1461–1481.
[73] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large sample covariance matrices and high-dimensional data analysis*. Cambridge University Press Cambridge.
[74] YU, L., ZHAO, P. and ZHOU, W. (2024 (in press)). Testing the number of common factors by bootstrapped sample covariance matrix in high-dimensional factor models. *Journal of the American Statistical Association*.
[75] ZHANG, G., JIANG, D. and YAO, F. (2024). Covariance test and universal bootstrap by operator norm. *arXiv preprint arXiv:2412.20019*.
[76] ZHANG, L. (2006). Spectral Analysis of Large Dimensional Random Matrices. *Ph.D. Thesis, National University of Singapore*.

Supplementary material

In this supplement, we provide the proofs for the main results and some auxiliary lemmas. As indicated in Section 6 that our actual proof relies on two technical inputs. One is the detailed analysis of the Stieltjes transforms of the limiting ESD on local scales in some carefully chosen spectral domains. The other one is a finer control of the randomness of the quantities associated with ESD. Then, in Section S.1, we provide some preliminary results and some necessary technique results including averaged local laws. In Section S.2, we prove the averaged local laws near the edges. In Section S.3, we provide the asymptotic locations of the edge eigenvalues and prove the main results and other results related to our statistical applications. Finally, with the above theoretical foundation, the proofs of the main results in Sections 3 and 4 will be put in Sections S.4 and S.5, some auxiliary lemmas are proved in Section S.6.

APPENDIX S.1: SOME PRELIMINARY RESULTS

In this section, we introduce some preliminary results which will be used in the proofs. First, in Section S.1.1, we provide the properties of the asymptotic local laws m_{1n}, m_{2n} and m_n from Definition 6.1 and establish the averaged local laws. Third, in Section S.1.2, we examine the properties of the entries of D^2 and construct some probability events to which our arguments will be restricted. Finally, in Section S.1.3, we provide some useful lemmas and a short review of the extreme value theory.

S.1.1. Properties of asymptotic laws and averaged local laws

We start with introducing the properties of the asymptotic local laws as in Definition 6.1. Recall the definitions of the Stieltjes transforms of ESDs in (6.1). In practice, it is convenient to define the Green functions of Q and $\mathcal{Q}$ for $z = E + \mathrm{i}\eta \in \mathbb{C}_+$,

$$G(z) = (Q - zI)^{-1} \in \mathbb{R}^{p\times p}, \quad \mathcal{G}(z) = (\mathcal{Q} - zI)^{-1} \in \mathbb{R}^{n\times n}. \tag{S.1}$$

Then (6.1) can be rewritten as

$$m_Q = \frac{1}{p}\operatorname{tr} G(z), \quad m_{\mathcal{Q}} = \frac{1}{n}\operatorname{tr} \mathcal{G}(z).$$

Before we proceed ahead, we revisit the definition of the systems of equations in Definition 6.1. Thanks to Theorem 6.2, it is easy to see that the study of the systems of equations in (6.2) can be reduced to the analysis of

$$F_n(m_{1n}(z), z) = 0, \quad z \in \mathbb{C}_+, \tag{S.2}$$

where $F_n(\cdot,\cdot)$ are defined as follows

$$F_n(m_{1n}(z), z) = \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-z + \frac{\sigma_i}{n}\sum_{j=1}^{n} \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}} - m_{1n}(z). \tag{S.3}$$

In the actual proof, the conditional and unconditional version in (6.2), (6.3) are both useful in their own aspects. To be more specific, the conditional version is more powerful when ξ^2 has unbounded support whereas the unconditional version is more convenient when ξ^2 has bounded support. Moreover, to avoid the singularity in the definitions of the systems of equations, we introduce some additional assumption on Σ which will be used when the multiplier ξ^2 has bounded support in the sense of (ii) of Assumption 2.2. Such an assumption has been frequently used in the random matrix theory literature, for example, see [9, 26, 27, 28, 32, 51, 55]. Recall the notations $m_{2n,c}$ and L_+ in Remark 6.3.

ASSUMPTION S.1.1. When (ii) of Assumption 2.2 holds, for Σ satisfying Assumption 2.4, we assume that for some constant $\tau > 0$

$$\min_{1 \leqslant i \leqslant p} |1 + \sigma_i m_{2n,c}(L_+)| \geqslant \tau.$$

We now define the sets of spectral parameters as follows. For ξ^2 with unbounded support as in Case (i) of Assumption 2.2, for ϑ_1 defined in (6.5) and d_1 defined as

$$d_1 := \begin{cases} n^{1/\alpha - \epsilon}, & \text{if (2.3) holds;} \\ 1, & \text{if (2.4) holds,} \end{cases} \tag{S.4}$$

we denote for sufficiently large constant $\mathtt{C} > 0$ that

$$\mathbf{D}_u \equiv \mathbf{D}_u(\mathtt{C}) := \left\{ z = E + \mathrm{i}\eta \in \mathbb{C}_+ : |E - \vartheta_1| \leqslant \mathtt{C} d_1,\ n^{-2/3} \leq \eta \leq \mathtt{C}\vartheta_1 \right\}. \tag{S.5}$$

For ξ^2 with bounded support as in Case (ii) of Assumption 2.2, for some sufficiently small constants $\mathtt{c},\ \epsilon_d > 0$, we denote (recall L_+ in (6.4))

$$\mathbf{D}_b \equiv \mathbf{D}_b(\mathtt{c}) := \left\{ z = E + \mathrm{i}\eta \in \mathbb{C}_+ : L_+ - \mathtt{c} \leq E \leq L_+ + \mathtt{c},\ n^{-1/2-\epsilon_d} \leq \eta \leq n^{-1/(d+1)+\epsilon_d} \right\}. \tag{S.6}$$

Throughout the paper, we will frequently use the minors of a matrix. For the data matrix Y in (2.1), denote the index set $\mathcal{I} = \{1, \dots, n\}$. Given an index set $\mathcal{T} \subset \mathcal{I}$, we introduce the notation $Y^{(\mathcal{T})}$ to denote the $p \times (n - |\mathcal{T}|)$ minor of Y obtained from removing all the ith columns of Y for $i \in \mathcal{T}$ and keep the original indices of Y. In particular, $Y^{(\emptyset)} = Y$. For convenience, we briefly write $(\{i\})$, $(\{i,j\})$ and $\{i,j\} \cup \mathcal{T}$ as (i), (i,j) and $(ij\mathcal{T})$ respectively. Correspondingly, we denote their sample covariance matrices and resolvents as

$$Q^{(\mathcal{T})} = (Y^{(\mathcal{T})})(Y^{(\mathcal{T})})^*,\ \mathcal{Q}^{(\mathcal{T})} = (Y^{(\mathcal{T})})^*(Y^{(\mathcal{T})}). \tag{S.7}$$

and

$$G^{(\mathcal{T})}(z) = (Q^{(\mathcal{T})} - zI)^{-1}, \quad \mathcal{G}^{(\mathcal{T})}(z) = (\mathcal{Q}^{(\mathcal{T})} - zI)^{-1}. \tag{S.8}$$

Similar to (6.1) and Definition 6.1, we can define $m_Q^{(\mathcal{T})}(z)$, $m_{\mathcal{Q}}^{(\mathcal{T})}(z)$, $m_{1n}^{(\mathcal{T})}(z)$, $m_{2n}^{(\mathcal{T})}(z)$ and $m_n^{(\mathcal{T})}(z)$ by removing $\mathbf{y}_i, i \in \mathcal{T}$ or $\xi_i^2, i \in \mathcal{T}$.

We begin with the summary of the results when ξ^2 has unbounded support as in Case (i) of Assumption 2.2. The proofs will be deferred to Section S.6.1. Conditional on some probability event, we provide some useful deterministic estimates for m_{1n}, m_{2n} and $m_n(z)$ on the above concerned spectral domain (S.5). Denote the control parameter e as follows

$$e := \begin{cases} \frac{\log n}{n^{1/\alpha}}, & \text{if (2.3) holds;} \\ \frac{1}{\log^{1/\beta} n}, & \text{if (2.4) holds.} \end{cases} \tag{S.9}$$

LEMMA S.1.2. *Suppose Assumptions 2.1, 2.4 and (i) of Assumption 2.2 hold. For any fixed realization $\{\xi_i^2\} \in \Omega$ where $\Omega \equiv \Omega_n$ is some probability event that $\mathbb{P}(\Omega) = 1 - \mathrm{o}(1)$, we have*

1. *For $z \in \mathbf{D}_u$, we have that for some constants $C_1, C_2 > 0$*
$$\operatorname{Re} m_{1n}(z) \asymp -E^{-1}, \; C_1 \eta E^{-2} \leqslant \operatorname{Im} m_{1n}(z) \leqslant C_2 \eta E^{-1}.$$
2. *When $|E - \mu_1| \leqslant C d_1$ for some sufficiently large constant $C > 0$, let $m_{1n}(E) = \lim_{\eta \downarrow 0} m_{1n}(E + \mathrm{i}\eta)$, then we have that*
$$m_{1n}(E) \asymp -E^{-1}.$$
3. *For $z \in \mathbf{D}_u$ and e defined in (S.9), we have that*
$$|m_{2n}(z)| = \mathrm{O}(e), \quad |m_n(z)| = \mathrm{O}(E^{-1}),$$
$$\operatorname{Im} m_{2n}(z) = \mathrm{O}(\eta E^{-1}), \quad \operatorname{Im} m_n(z) = \mathrm{O}(\eta E^{-2}).$$

REMARK S.1.3. The above lemma provides some controls for the Stieltjes transforms. Four remarks are in order. First, the construction of the probability event Ω will be given in Section S.1.2. Second, by a discussion similar to (S.8), conditional on Ω, we can replace μ_1 with $\varphi \xi_{(1)}^2$. Third, The above results hold when we replace m_{1n}, m_{2n} and m_n with $m_{1n}^{(\mathcal{T})}, m_{2n}^{(\mathcal{T})}$ and $m_n^{(\mathcal{T})}$ for any finite $\mathcal{T}$. Fourth, Lemma S.1.2 also implies the existence of ϑ_1 defined in (3.2).

Then we state the results when ξ^2 has bounded support as in Case (ii) of Assumption 2.2. As mentioned in Remark 6.3, for the bounded support case, it will be more convenient to use both the conditional and unconditional systems. For the conditional setting, when restricted to Ω, we denote the rightmost edge of ρ as $\widehat{L}_+$. Moreover, parallel to (3.5), we introduce the following quantities

$$\widehat{\mathsf{s}}_1 := \frac{1}{n} \sum_{j=1}^n \frac{l^2 \xi_j^4}{(l - \xi_j^2)^2}, \quad \widehat{\mathsf{s}}_2 := \frac{1}{n} \sum_{j=1}^n \frac{l \xi_j^2}{l - \xi_j^2}, \quad \widehat{\mathsf{s}}_3 := \frac{1}{p} \sum_{i=1}^p \frac{\sigma_i^2 \widehat{\mathsf{s}}_1}{(\widehat{L}_+ - \sigma_i \widehat{\mathsf{s}}_2)^2}, \tag{S.10}$$

$$\widehat{\mathsf{s}}_4 := \frac{1}{n} \sum_{i=1}^p \frac{\sigma_i}{(-\widehat{L}_+ + \sigma_i \widehat{\mathsf{s}}_2)^2}. \tag{S.11}$$

Furthermore, we need the following spectral parameter set

$$\mathbf{D}_b' = \left\{ z \in \mathbf{D}_b : |1 + \xi_j^2 m_{1n,c}(z)| > \frac{1}{2} n^{-1/(d+1) - \epsilon_d}, \quad \text{for all } 2 \leqslant j \leqslant n \right\}. \tag{S.12}$$

REMARK S.1.4. We will see from (S.22) and (S.27) that with probability $1-\mathrm{o}_{\mathbb{P}}(1)$, $\lambda_1+\mathrm{i}\eta_0\in\mathbf{D}_b'$.

LEMMA S.1.5. *Suppose Assumptions 2.1, 2.4, S.1.1 and (ii) of Assumption 2.2 hold. Then for any fixed realization $\{\xi_i^2\}\in\Omega$ where $\Omega\equiv\Omega_n$ is some probability event that $\mathbb{P}(\Omega)=1-\mathrm{o}(1)$, for sufficiently large n, we have that*

(a). *If $d>1$ and $\phi^{-1}>\widehat{\mathsf{s}}_3$, $\widehat{L}_+$ can be expressed explicitly by the following equation*

$$1=\frac{1}{n}\sum_i\frac{-l\sigma_i}{(-\widehat{L}_++\sigma_i\widehat{\mathsf{s}}_2)}. \tag{S.13}$$

Moreover, for any $0\le\kappa\le\widehat{L}_+$,

$$\rho(\widehat{L}_+-\kappa)\asymp\kappa^d, \tag{S.14}$$

Moreover, for some sufficiently small constant $\epsilon>0$

$$\mathsf{s}_k=\widehat{\mathsf{s}}_k+\mathrm{O}(n^{-1/2+\epsilon}), k=1,2,3,4;\ L_+=\widehat{L}_++\mathrm{O}(n^{-1/2+\epsilon}). \tag{S.15}$$

In addition, let $z=\widehat{L}_+-\kappa+\mathrm{i}\eta\in\mathbf{D}_b$, then

$$m_{1n}(\widehat{L}_+)-m_{1n}(z)=\frac{\widehat{\mathsf{s}}_4}{(1-\phi\widehat{\mathsf{s}}_3)}\left(\widehat{L}_+-z\right)+\mathrm{O}((\log n)(\kappa+\eta)^{\min\{d,2\}}). \tag{S.16}$$

Similarly, for any $z,z'\in\mathbf{D}_b$, we have

$$m_{1n}(z)-m_{1n}(z')=\frac{\widehat{\mathsf{s}}_4}{(1-\phi\widehat{\mathsf{s}}_3)}(z-z')+\mathrm{O}((\log n)(n^{-1/(d+1)})^{\min\{d-1,1\}}|z-z'|). \tag{S.17}$$

Finally, for $z\in\mathbf{D}_b'$ in (S.12), we have that

$$\operatorname{Im} m_{1n}(z)=\mathrm{O}\left(\max\left\{\eta,\frac{1}{n\eta}\right\}\right),\ \operatorname{Im} m_n(z)=\mathrm{O}\left(\max\left\{\eta,\frac{1}{n\eta}\right\}\right). \tag{S.18}$$

Moreover, for z_0 defined in (6.7), we have that

$$\operatorname{Im} m_{1n}(z_0)\asymp n^{-1/2},\ \operatorname{Im} m_n(z_0)\asymp n^{-1/2}. \tag{S.19}$$

and for $z=E+\mathrm{i}\eta_0\in\mathbf{D}_b'$ in (S.12), we have that

$$\operatorname{Im} m_{1n}(z)\asymp\eta_0,\ \operatorname{Im} m_n(z)\asymp\eta_0, \tag{S.20}$$

if $|z-z_0|\geqslant Cn^{-1/2+3\epsilon_d}$ for some constant $C>0$.

(b). *If $-1<d\leqslant1$ or $d>1$ and $\phi^{-1}<\widehat{\mathsf{s}}_3$, we have that for some fixed constant $\tau>0$ and all $1\leqslant i\leqslant n$*

$$\left|1+\xi_i^2m_{1n}(\widehat{L}_+)\right|\geqslant\tau. \tag{S.21}$$

Moreover, we have that for any $0\le\kappa\le\widehat{L}_+$,

$$\rho(\widehat{L}_+-\kappa)\asymp\kappa^{1/2}. \tag{S.22}$$

Equivalently, for some constant $\gamma>0$, we have for $\kappa\downarrow0$

$$\rho(\widehat{L}_+-\kappa)=\frac{1}{\pi}\gamma^{3/2}\sqrt{\kappa}+\mathrm{O}(\kappa). \tag{S.23}$$

Finally, the results of (a) an (b) still hold unconditionally when m_{1n} is replaced by $m_{1n,c}$, ρ is replaced by $\widetilde{\rho}$ and $\widehat{L}_+$ is replaced by L_+ as in Remark 6.3 where $\widehat{\mathsf{s}}_3$ and $\widehat{\mathsf{s}}_4$ should be replaced by s_3 and s_4 as in (3.5).

REMARK S.1.6. Using discussions similar to the paragraphs around equation (5.1) of [52], by (S.16), (S.27) and the fact $m_{1n}(\widehat{L}_+) = -l^{-1}$, we see that (6.7) has at least one solution.

Then we provide the results of the averaged local laws. Throughout the paper, we will consistently use the notion of *stochastic domination* to systematize the statements of the form "ξ is bounded by ζ with high probability up to a small power of n."

DEFINITION S.1.7 (Stochastic domination). *(i) Let*

$$\xi = \left(\xi^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right), \quad \zeta = \left(\zeta^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right),$$

be two families of nonnegative random variables, where $U^{(n)}$ is a possibly n-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any fixed (small) $\epsilon > 0$ and (large) $D > 0$,

$$\sup_{u \in U^{(n)}} \mathbb{P}\left(\xi^{(n)}(u) > n^{\epsilon}\zeta^{(n)}(u)\right) \le n^{-D}$$

for large enough $n \ge n_0(\epsilon, D)$, and we shall use the notation $\xi \prec \zeta$. Throughout this paper, the stochastic domination will always be uniform in all parameters that are not explicitly fixed, such as the matrix indices and the spectral parameter z. If for some complex family ξ we have $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = \mathrm{O}_{\prec}(\zeta)$.

(ii) We say an event Ξ holds with high probability if for any constant $D > 0$, $\mathbb{P}(\Xi) \ge 1 - n^{-D}$ for large enough n.

Similar to [17, 27, 64, 71], instead of working directly with m_Q and $m_{\mathcal{Q}}$ in (6.1), it is more convenient to study the following quantities

$$m_1(z) = \frac{1}{n}\operatorname{tr}\left(G(z)\Sigma\right), \quad m_2(z) = \frac{1}{n}\sum_{i=1}^{n} \xi_i^2 \mathcal{G}_{ii}(z). \tag{S.24}$$

Analogously, using the minors in (S.7), we can define $m_1^{(\mathcal{T})}(z)$ and $m_2^{(\mathcal{T})}(z)$. The following Theorem S.1.8 summarizes the averaged local laws for unbounded ξ^2 which will be used in our proof for the main results. Its proof can be found in Section S.2.1.

THEOREM S.1.8 (Averaged local laws for unbounded support ξ^2). *Suppose Assumptions 2.1, 2.4 and (i) of Assumption 2.2 hold. For any fixed realization $\{\xi_i^2\} \in \Omega$ where $\Omega \equiv \Omega_n$ is introduced in Lemma S.1.2, let $m_1^{(1)}(z)$ and $m_{1n}^{(1)}(z)$ be defined by removing the column or entries associated with $\xi_{(1)}^2$. We have that the followings hold true uniformly for $z \in \mathbf{D}_u$ in (S.5)*

1. *If* Case (a) *of (i) of Assumption 2.2 holds, we have that*
$$m_1^{(1)}(z) = m_{1n}^{(1)}(z) + \mathrm{O}_{\prec}\left(n^{-1/2-2/\alpha}\right).$$
2. *If* Case (b) *of (i) of Assumption 2.2 holds, we have that*
$$m_1^{(1)}(z) = m_{1n}^{(1)}(z) + \mathrm{O}_{\prec}\left(n^{-1/2}\right).$$

Then we provide the averaged local laws for bounded ξ^2. Recall m_Q defined in (6.1).

THEOREM S.1.9 (Averaged local laws for bounded support ξ^2). *Suppose Assumptions 2.1, 2.4, S.1.1 and (ii) of Assumption 2.2 hold. When $d>1$ and $\phi^{-1} > \mathsf{s}_3$, for any fixed realization $\{\xi_i^2\} \in \Omega$ where $\Omega \equiv \Omega_n$ is introduced in Lemma S.1.5, for $\eta_0 = n^{-1/2-\epsilon_d}$ defined in (6.7), we have that the followings hold true uniformly for $z \in \mathbf{D}_b'$ in (S.12)*

$$|m_{1n}(z) - m_{1n,c}(z)| \leqslant n^{-1/2+\epsilon_d}, \ |m_1(z) - m_{1n}(z)| = \mathrm{O}_{\prec}\left((n\eta_0)^{-1}\right),$$

and

$$|m_n(z) - m_{n,c}(z)| \leqslant n^{-1/2+\epsilon_d}, \ |m_Q(z) - m_n(z)| = \mathrm{O}_{\prec}\left((n\eta_0)^{-1}\right).$$

S.1.2. Characterization of "good configurations"

In this subsection, independent of Section S.1.1, we define some probability events which are some "good configurations" for the first few largest eigenvalues of D^2. Our proofs will be restricted on these probability events. In fact, as will be seen in Lemma S.1.12, under Assumption 2.2, these probability events hold with high probability when n is sufficiently large.

Recall Assumption 2.2 and

$$D^2 = \mathrm{diag}\left\{\xi_1^2, \cdots, \xi_n^2\right\}.$$

Moreover, we define the order statistics of $\{\xi_i^2\}$ as

$$\xi_{(1)}^2 \geqslant \xi_{(2)}^2 \geqslant \cdots \geqslant \xi_{(n)}^2.$$

In what follows, we define these probability events according to the various assumptions of $\{\xi_i^2\}$ in (2.3)–(2.5).

DEFINITION S.1.10. *Denote $\Omega \equiv \Omega_n$ be the event on $\{\xi_i^2\}$ so that the following conditions hold:*

(a). Unbounded support with polynomial decay. *When $\{\xi_i^2\}$ has unbounded support with polynomial decay tail as in (2.3), we assume that for all $\epsilon \in (0, 1/\alpha)$, $b \in (1/2, 1]$ and some constants $C, c > 1$, the following holds on Ω*

$$\begin{aligned}
&\xi_{(1)}^2 - \xi_{(2)}^2 \geq C^{-1} n^{1/\alpha} \log^{-1} n, \\
&C^{-1} n^{1/\alpha} \log^{-1} n \leq \xi_{(1)}^2 \leq C n^{1/\alpha} \log n, \\
&\xi_{(i)}^2 - \xi_{(i+1)}^2 \geq C^{-1} n^{\epsilon} \log^{-1} n, \ 1 \leqslant i < \sqrt{n}, \\
&\xi_{(1)}^2 - \xi_{(\lceil n^b \rceil)}^2 \geq c^{-1} n^{1/\alpha} \log^{-1} n, \\
&\frac{1}{n} \sum_{i=1}^n \xi_i^2 < \infty.
\end{aligned} \tag{S.25}$$

(b). Unbounded support with exponential decay. *When $\{\xi_i^2\}$ has unbounded support with polynomial decay tail as in (2.4), we assume that for some constant $C > 1$, the following holds on Ω*

$$\begin{aligned}
&\xi_{(1)}^2 - \xi_{(2)}^2 \geq C^{-1} \log^{1/\beta} n, \\
&C^{-1} \log^{1/\beta} n \leq \xi_{(1)}^2 \leq C \log^{1/\beta} n, \\
&\frac{1}{n} \sum_{i=1}^n \xi_i^2 < \infty.
\end{aligned} \tag{S.26}$$

(c). Bounded support with $d>1$. *When $\{\xi_i^2\}$ has bounded support satisfying (2.5) with $d>1$, we assume that for some sufficiently small constant $\epsilon>0$, $\epsilon_d<1/8(1/2-1/(d+1))$, $0<b\leqslant 1$ and $0<C_l<l$, the following holds on Ω*

$$
\begin{aligned}
& n^{-1/(d+1)-\epsilon_d} < l-\xi^2_{(1)} < n^{-1/(d+1)}\log n, \\
& \xi^2_{(1)}-\xi^2_{(2)} > n^{-1/(d+1)-\epsilon_d}, \\
& l-\xi^2_{(\lfloor bn \rfloor)} > C_l, \\
& \frac{1}{n}\sum_{i=1}^n \xi_i^2 \leq l, \\
& \left| \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^2}{1+\xi_i^2 m_{1n,c}(z)} - \int \frac{t}{1+t m_{1n,c}(z)} \mathrm{d}F(t) \right| \leq \frac{Cn^{\epsilon}}{\sqrt{n}}, \text{ for } z\in \mathbf{D}_b,
\end{aligned} \tag{S.27}
$$

where we recall that $F(t)$ is the distribution of ξ^2 and $C>0$ is some generic constant.

REMARK S.1.11. Two remarks are in order. First, on the event Ω, for the unbounded support case, according to (a) and (b), we see that the first few largest ξ_i^2 are divergent and well separated from each other. Second, for the bounded support case, we only provide the results for $d>1$ in (c). Nevertheless, it is easy to see that similar results can be obtained for $-1<d\leqslant 1$.

The following lemma shows that under Assumption 2.2, the probability event Ω happens with high probability in all the four settings. The proof will be given in Section S.6.2.

LEMMA S.1.12. *Let Ω be the events defined in Definition S.1.10, suppose Assumption 2.2 holds, we then have that when n is sufficiently large*

$$\mathbb{P}(\Omega)=1-\mathrm{O}(\log^{-D} n),$$

for some constant $D>0$.

S.1.3. Some useful lemmas and a summary of extreme value theory

In this subsection, we first provide some technical lemmas which will be used in our proof. The following resolvent identities play an important role in our proof. Recall the resolvents defined in (S.1) and the minors defined in (S.7).

LEMMA S.1.13 (Resolvent identities). *Let $\{\mathbf{y}_i\}\subset \mathbb{R}^p$ be the columns of Y as in (2.1), then we have that*

$$
\begin{aligned}
\mathcal{G}_{ii}(z) &= -\frac{1}{z+z\mathbf{y}_i^* G^{(i)}(z)\mathbf{y}_i}, \\
\mathcal{G}_{ij}(z) &= z\mathcal{G}_{ii}(z)\mathcal{G}_{jj}^{(i)}(z)\mathbf{y}_i^* G^{(ij)}(z)\mathbf{y}_j \quad i\neq j, \\
\mathcal{G}_{ij}(z) &= \mathcal{G}_{ij}^{(k)}(z)+\frac{\mathcal{G}_{ik}(z)\mathcal{G}_{kj}(z)}{\mathcal{G}_{kk}(z)} \quad i,j\neq k.
\end{aligned}
$$

PROOF. The proof is straightforward using Schur's complement formula; for example see [65, Lemma 2.3]. □

LEMMA S.1.14 (Some useful matrix identities). *For any finite subset $\mathcal{T} \subset \{1, \ldots, n\}$, we have that*

$$\left\| G^{(\mathcal{T})} \Sigma^{1/2} \right\|_F^2 = \eta^{-1} \operatorname{Im} \operatorname{Tr} \left(G^{(\mathcal{T})} \Sigma \right). \tag{S.28}$$

Moreover, we have that

$$\begin{aligned} \left| \operatorname{Tr}(\mathcal{G}^{(i)} - \mathcal{G}) \right| &\leq \eta^{-1}, \\ \left| \operatorname{Tr}(G^{(i)} - G) \right| &\leq |z|^{-1} + \eta^{-1}, \\ \left| \operatorname{Im} \operatorname{Tr}(G^{(i)} - \mathcal{G}) \right| &\leq \eta |z|^{-2} + \eta^{-1}. \end{aligned} \tag{S.29}$$

PROOF. Due to similarity, we focus our discussion on the separable covariance i.i.d. data, i.e., Case (2) of Assumption 2.1. In fact, it is easier to handle Case (1) since Σ can be always assumed to be diagonal.

We start with the proof of (S.28). Recall (S.8). We can write

$$G^{(\mathcal{T})} = \left(\Sigma^{1/2} X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} \Sigma^{1/2} - z \right)^{-1}.$$

Let the spectral decomposition $\Sigma = U \Lambda U^*$. Observe that

$$\begin{aligned} \|G^{(\mathcal{T})} \Sigma^{1/2}\|_F^2 &= \operatorname{Tr} \left(\left(\Sigma^{1/2} X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} \Sigma^{1/2} - z \right)^{-1} U \Lambda U^* \left(\Sigma^{1/2} X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} \Sigma^{1/2} - \bar{z} \right)^{-1} \right) \\ &= \operatorname{Tr} \left(U \left(\Lambda^{1/2} U^* X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} U \Lambda^{1/2} - z \right)^{-1} U^* U \Lambda U^* U \left(\Lambda^{1/2} U^* X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} U \Lambda^{1/2} - \bar{z} \right)^{-1} U^* \right) \\ &= \left\| \left(\Lambda^{1/2} U^* X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} U \Lambda^{1/2} - z \right)^{-1} \Lambda^{1/2} \right\|_F^2 \\ &= \eta^{-1} \operatorname{Im} \operatorname{Tr} \left[\left(\Lambda^{1/2} U^* X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} U \Lambda^{1/2} - z \right)^{-1} \Lambda \right], \\ &= \eta^{-1} \operatorname{Im} \operatorname{Tr} \left[U^* U \left(\Lambda^{1/2} U^* X^{(\mathcal{T})} D^2 X^{(\mathcal{T})} U \Lambda^{1/2} - z \right)^{-1} U^* U \Lambda \right], \\ &= \eta^{-1} \operatorname{Im} \operatorname{Tr} \left(G^{(\mathcal{T})} \Sigma \right), \end{aligned}$$

where in the fourth step we used Ward's identity (see the equation below (4.42) of [22]).

Second, the proof of (S.29) follows from the definitions of the resolvents; see [26, Lemma A.4] and the proof of [9, Lemma 4.6] for more detail.

□

LEMMA S.1.15 (Large deviation bounds). *Let $\mathbf{u} = (u_1, u_2, \cdots, u_p)^*, \widetilde{\mathbf{u}} = (\widetilde{u}_1, \widetilde{u}_2, \cdots, \widetilde{u}_p)^* \in \mathbb{R}^p$ be two real independent random vectors. Moreover, let A be a $p \times p$ matrix independent of the above vectors. Suppose the entries of the random vectors are centered i.i.d. random variables with variance n^{-1} and $\mathbb{E}|\sqrt{n} v_i|^k \leqslant C_k$, where $v_i = u_i, \widetilde{u}_i, 1 \leqslant i \leqslant p$, then we have that*

$$|\widetilde{\mathbf{u}}^* \mathbf{u}| \prec \sqrt{\frac{\|\mathbf{u}\|^2}{n}}, \ \ |\mathbf{u}^* A \widetilde{\mathbf{u}}| \prec \frac{1}{n} \|A\|_F, \ \ |\mathbf{u}^* A \mathbf{u} - \frac{1}{n} \operatorname{Tr} A| \prec \frac{1}{n} \|A\|_F.$$

PROOF. The proof can be found in Lemma 3.4 of [65] or Lemma 5.6 of [71]. □

In what follows, we provide a mini-review of the extreme value theory for a sequence of i.i.d. random variables following [42]. For more systematic treatments, we refer to the monographs [10, 16, 40, 67].

LEMMA S.1.16 (Fisher-Tippett-Gnedenko Theorem). *Let $\{x_i^2\}$ be a sequence of i.i.d. random variables and denote $M_n := x_{(1)}^2$ as the largest order statistic.*

1. *If there exist some constants $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$ and some non-degenerate* cdf *G such that $\alpha_n^{-1}(M_n - \beta_n)$ converges in distribution to G, then G belongs to the type of one of the following three* cdfs:

$$\textit{Gumbel}: \ G_0(x) = \exp(-e^{-x}), \ x \in \mathbb{R},$$
$$\textit{Fréchet}: \ G_{1,\alpha}(x) = \exp(-x^{-\alpha}), \ x \geqslant 0, \ \alpha > 0,$$
$$\textit{Weibull}: \ G_{2,\alpha}(x) = \exp(-|x|^{\alpha}), \ x \leqslant 0, \ \alpha > 0.$$

2. *Recall (3.1). First, if $\{x_i\}$ satisfies (2.3), we have that*

$$\frac{M_n}{b_n} \overset{d}{\Rightarrow} G_{1,\alpha},$$

Moreover, if we futher assume $\lim_{x\uparrow\infty} L(x) = \mathsf{C}$ for some constant $\mathsf{C} > 0$, then $b_n = (\mathsf{C}n)^{1/\alpha}$. Second, if $\{x_i\}$ satisfies (2.4) and (2.8), we have that

$$\mathsf{g}'(b_n)(M_n - b_n) \overset{d}{\Rightarrow} G_0.$$

Finally, if (2.5) holds, recall $\mathfrak{b}$ in (2.6), we have that

$$(\mathfrak{b}n)^{1/(d+1)}(M_n - l) \overset{d}{\Rightarrow} G_{2,d+1}.$$

PROOF. The proof can be found in the standard textbook or review article regarding extreme value theory. For example, see [42] and [10]. □

APPENDIX S.2: PROOF OF AVERAGED LOCAL LAWS

In this section, we prove the local laws Theorems S.1.8 and S.1.9.

S.2.1. Unbounded support setting: proof of Theorem S.1.8

In this section, we will prove Theorem S.1.8. Due to similarity, we focus on the proof of part 1 and briefly discuss that of part 2. The proof contains two steps. In the first step, we will establish the results for the results of Q outside the bulk of the spectrum on the domain $\widetilde{\mathbf{D}}_u$ denoted as follows

$$\widetilde{\mathbf{D}}_u \equiv \widetilde{\mathbf{D}}_u(\mathtt{C}) := \left\{ z = E + \mathrm{i}\eta : 0 < E - \vartheta_1 \leqslant \mathtt{C} d_1, \ n^{-2/3} \leq \eta \leq \mathtt{C}\vartheta_1 \right\}. \tag{S.1}$$

That is, we will establish the following proposition.

PROPOSITION S.2.1. *Under the assumptions of Theorem S.1.8, the following results hold uniformly on the spectral domain $\widetilde{\mathbf{D}}_u$ in (S.1) when conditional on the event Ω in Lemma S.1.12.*

(1). *If* Case (a) *of (i) of Assumption 2.2 holds, we have that*

$$\mathcal{G}_{ij}(z) = -\frac{\delta_{ij}}{z\big(1 + m_{1n}(z)\xi_i^2\big)} + \mathrm{O}_{\prec}\left(n^{-1/2-1/\alpha}\right), \tag{S.2}$$

where δ_{ij} is the Dirac delta function so that $\delta_{ij}=1$ when $i=j$ and $\delta_{ij}=0$ when $i\neq j$. Moreover, we have that

$$m_1(z)=m_{1n}(z)+\mathrm{O}_{\prec}\left(n^{-1/2-2/\alpha}\right),\ m_2(z)=m_{2n}(z)+\mathrm{O}_{\prec}\left(n^{-1/2-1/\alpha}\right),$$

and

$$m_Q(z)=m_n(z)+\mathrm{O}_{\prec}\left(n^{-1/2-2/\alpha}\right). \tag{S.3}$$

(2). *If* Case (b) *of (i) of Assumption 2.2 holds, we have that the results in part (1) hold by setting $\alpha=\infty$.*

Once Proposition S.2.1 is proved, we can quantify the rough locations of the eigenvalues of Q as summarized in the following lemma.

LEMMA S.2.2. *Suppose Assumptions 2.1, 2.4 and (i) of Assumption 2.2 hold. For some sufficiently large constant $C>0$, with high probability, for any fixed realization $\{\xi_i^2\}\in\Omega$ where $\Omega\equiv\Omega_n$ is introduced in Lemma S.1.12, for all $1\leqslant i\leqslant\min\{p,n\}$, we have that*

$$\lambda_i(Q)\notin(\vartheta_1, Cn^{1/\alpha}\log n),\ \textit{if}\text{ Case (i)-a of Assumption 2.2 holds,} \tag{S.4}$$

and

$$\lambda_i(Q)\notin(\vartheta_1, C\log^{1/\beta} n),\ \textit{if}\text{ Case (i)-b of Assumption 2.2 holds.} \tag{S.5}$$

PROOF. Due to similarity, we focus our arguments on (S.4). We prove the results by contradiction. Assume there is an eigenvalue of Q lies in the interval as in (S.4), denote as $\widehat{\lambda}$. Let $z=\widehat{\lambda}+\mathrm{i}n^{-2/3}$. Since $z\in\widetilde{\mathbf{D}}_u\subset\mathbf{D}_u$ as in (S.5), by Lemma S.1.2, we obtain $\operatorname{Im} m_n(z)=\eta\widehat{\lambda}^{-2}$. According to (S.8) and (S.25), we have that on the event Ω

$$\vartheta_1\gtrsim n^{1/\alpha}\log^{-1} n. \tag{S.6}$$

Together with (S.3), we readily see that

$$\begin{aligned}\operatorname{Im} m_Q(z)&=\operatorname{Im} m_n(z)+\operatorname{Im}(m_Q(z)-m_n(z))\\&\prec n^{-2/\alpha-2/3}+n^{-1/2-2/\alpha}\prec n^{-1/2-2/\alpha}.\end{aligned} \tag{S.7}$$

On the other hand, we have

$$\operatorname{Im} m_Q(z)=\frac{1}{n}\sum_i\frac{\eta}{(\lambda_i-\widehat{\lambda})^2+\eta^2}\geq\frac{1}{n\eta}=n^{-1/3},$$

which contradicts (S.7). Therefore, there is no eigenvalue in this interval. Similarly, we can prove (S.5). The only difference is that (S.6) should be replaced by $\vartheta_1\gtrsim\log^{1/\beta} n$ according to (S.26) so that the error rate in (S.7) should be updated to $n^{-1/2}$. This completes the proof. □

Armed with the above lemma, we can proceed to the second step to conclude the proof of Theorem S.1.8. In what follows, we first provide the proof of Proposition S.2.1 in Section S.2.1.1. After that, we prove Theorem S.1.8 in Section S.2.1.2.

S.2.1.1. Proof of Proposition S.2.1

We first prepare two lemmas. The first one is to establish Proposition S.2.1 for large scale of η.

LEMMA S.2.3 (Average local law for large η). *Proposition S.2.1 holds when* $\eta = \mathtt{C}\vartheta_1$.

PROOF. In the sequel, without loss of generality, we assume that $\xi_1^2 \geqslant \xi_2^2 \geqslant \cdots \geqslant \xi_n^2$.
Note that according to (S.8), (S.25) and the definition of d_1 in (S.4), on the event Ω, we have that

$$E \asymp \xi_1^2. \tag{S.8}$$

When $\eta = \mathtt{C}E$, we have $\max\left\{\|G^{(\mathcal{T})}\|, \|\mathcal{G}^{(\mathcal{T})}\|\right\} \leq \eta^{-1} = \mathtt{C}^{-1}E^{-1}$ for any finite $\mathcal{T} \subset \{1, \ldots, n\}$ by definition. The main idea is to explore the relation of m_1 and m_2 using Lemma S.1.13. We start with m_2. By Lemma S.1.13 and definition of m_2 in (S.24), we have

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i^2}{-z - z\mathbf{y}_i^* G^{(i)}\mathbf{y}_i} = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i^2}{-z(1 + \xi_i^2 n^{-1}\mathrm{tr}G^{(i)}\Sigma + Z_i)}, \tag{S.9}$$

$$Z_i = \mathbf{y}_i^* G^{(i)}\mathbf{y}_i - \xi_i^2 n^{-1}\mathrm{tr}G^{(i)}\Sigma.$$

As $\mathbf{y}_i$ is independent of $G^{(i)}$, by (1) of Lemma S.1.15, we see that

$$Z_i \prec \frac{\xi_i^2}{n}\|G^{(i)}\Sigma\|_F \leq \frac{\xi_i^2}{n}\|G^{(i)}\|\|\Sigma\|_F \prec \frac{\xi_i^2}{\sqrt{n}\eta}. \tag{S.10}$$

Moreover, using the definition of m_1 in (S.24) and the second resolvent identity, we readily obtain that for some constant $C > 0$

$$\frac{1}{n}\mathrm{tr}(G^{(i)}\Sigma) - m_1(z) = \frac{1}{n}\mathbf{y}_i^* G\Sigma G^{(i)}\mathbf{y}_i \leq C\frac{\xi_i^2}{n\eta^2}. \tag{S.11}$$

Moreover, by (S.8) and the form of η, we find that for some constant $C > 0$

$$|1 + \xi_i^2 m_1| \geq 1 - C\mathtt{C}^{-1} > 0,$$

when $\mathtt{C} > 0$ is chosen to be sufficiently large. Together with (S.9), we obtain

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i^2}{-z(1 + \xi_i^2 m_1 + \mathrm{O}_\prec(\frac{\xi_i^2}{\sqrt{n}\eta}))} = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i^2}{-z(1 + \xi_i^2 m_1)} + \mathrm{O}_\prec(n^{-1/2-1/\alpha}), \tag{S.12}$$

where we used (S.25).

Then we work with m_1. Decompose that

$$Q - zI = \sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^* + zm_2(z)\Sigma - z(I + m_2(z)\Sigma).$$

Applying resolvent expansion to the order one, we obtain that

$$G = -z^{-1}(I + m_2(z)\Sigma)^{-1} + z^{-1}G\left(\sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^* + zm_2(z)\Sigma\right)(I + m_2(z)\Sigma)^{-1}.$$

Furthermore, using Shernman-Morrison formula, we have that

$$G\mathbf{y}_i = \frac{G^{(i)}\mathbf{y}_i}{1 + \mathbf{y}_i^* G^{(i)}\mathbf{y}_i}. \tag{S.13}$$

Combining the above two identities and Lemma S.1.13, we can further write

$$
\begin{aligned}
G &= -z^{-1}(I+m_2(z)\Sigma)^{-1} + \left[z^{-1}\sum_{i=1}^{n} \frac{G^{(i)}(\mathbf{y}_i\mathbf{y}_i^* - n^{-1}\xi_i^2\Sigma)}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2(z)\Sigma)^{-1}\right] \\
&\quad + \left[z^{-1}\frac{1}{n}\sum_{i=1}^{n} \frac{(G^{(i)}-G)\xi_i^2\Sigma}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2(z)\Sigma)^{-1}\right] \qquad \text{(S.14)} \\
&:= -z^{-1}(I+m_2(z)\Sigma)^{-1} + R_1 + R_2.
\end{aligned}
$$

In what follows, we control the two error terms R_1, R_2. For R_1, we notice that

$$
\begin{aligned}
\frac{z}{n}\operatorname{tr}(R_1\Sigma) &= \frac{1}{n}\sum_i \operatorname{tr}\left(\frac{G^{(i)}(\mathbf{y}_i\mathbf{y}_i^* - n^{-1}\xi_i^2\Sigma)}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2^{(i)}\Sigma)^{-1}\Sigma\right) \\
&+ \frac{1}{n}\sum_i \operatorname{tr}\left(\frac{G^{(i)}(\mathbf{y}_i\mathbf{y}_i^* - n^{-1}\xi_i^2\Sigma)}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2\Sigma)^{-1}(m_2^{(i)}-m_2)\Sigma(I+m_2^{(i)}\Sigma)^{-1}\Sigma\right) \qquad \text{(S.15)} \\
&:= \mathtt{R}_{11} + \mathtt{R}_{12}.
\end{aligned}
$$

Since $\|\mathcal{G}^{(i)}\| \le \eta^{-1}$, using (S.8), (S.25), with high probability, we have that for some constant $C>0$,

$$
|m_2^{(i)}(z)| \le \frac{1}{n}\sum_{j\neq i}\xi_j^2|\mathcal{G}_{jj}^{(i)}| \le C\frac{\log^2 n}{n^{1/\alpha}}. \qquad \text{(S.16)}
$$

Moreover, according to (S.10), with high probability, when n is sufficiently large, we have that for some constant $C>0$

$$
|1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i| \asymp |1+\xi_i^2 n^{-1}\operatorname{tr}G^{(i)}\Sigma| \geqslant 1 - C\mathtt{C}^{-1} > 0, \qquad \text{(S.17)}
$$

whenever $\mathtt{C}$ is chosen sufficiently large. Consequently, for all i, we have that

$$
\begin{aligned}
\operatorname{tr}\Big(\frac{G^{(i)}(\mathbf{y}_i\mathbf{y}_i^* - n^{-1}\xi_i^2\Sigma)}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2^{(i)}\Sigma)^{-1}\Sigma\Big) &\asymp \operatorname{tr}\left(\xi_i^2 G^{(i)}(\mathbf{u}_i\mathbf{u}_i^* - n^{-1}I)(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\right) \\
&= \xi_i^2\left(\mathbf{u}_i^* G^{(i)}(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\mathbf{u}_i - n^{-1}\operatorname{tr}\left(G^{(i)}(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\right)\right) \\
&\prec \xi_i^2\frac{1}{\eta\sqrt{n}}, \qquad \text{(S.18)}
\end{aligned}
$$

where in the third step we used (1) of Lemma S.1.15. Together with (S.25) and (S.8), we find that

$$
\mathtt{R}_{11} \prec n^{-1/2-1/\alpha}.
$$

For $\mathtt{R}_{12}$, using the definition in (S.24) and the identity in Lemma S.1.13 and the definition of $\mathcal{G}^{(i)}$ (see (S.28) below), we see that

$$
m_2(z) - m_2^{(i)}(z) = \frac{1}{n}\sum_{j=1}^{n}\xi_j^2(\mathcal{G}_{jj}-\mathcal{G}_{jj}^{(i)}) = \frac{1}{n}\sum_{j\neq i}\xi_j^2\frac{\mathcal{G}_{ji}\mathcal{G}_{ij}}{\mathcal{G}_{ii}} + \frac{\xi_i^2(\mathcal{G}_{ii}-|z|^{-1})}{n}. \qquad \text{(S.19)}
$$

In addition, using Lemma S.1.13 and a discussion similar to (S.17), we conclude that

$$
\frac{1}{\mathcal{G}_{ii}(z)} = -z - z\mathbf{y}_i^* G^{(i)}\mathbf{y}_i \prec |z|.
$$

Moreover, by Lemmas S.1.13 and S.1.15, we have that

$$\mathcal{G}_{ij}(z)=z\mathcal{G}_{ii}(z)\mathcal{G}_{jj}^{(i)}(z)\mathbf{y}_i^*G^{(ij)}\mathbf{y}_j \prec |z|\eta^{-2}|\xi_i\xi_j|n^{-1}\|\mathcal{G}^{(ij)}\|_F \prec n^{-1/2}|z|\eta^{-3}|\xi_i\xi_j|, \quad i\neq j.$$

Combining the above bounds with (S.25), we see that

$$m_2(z)-m_2^{(i)}(z)\prec n^{-1-1/\alpha}.$$

Together with (S.17) and (S.18), we arrive at

$$\mathtt{R}_{12}\prec \frac{1}{\eta^2 n^{3/2}}.$$

Using the above bounds, we see that

$$\frac{z}{n}\operatorname{tr}(R_1\Sigma)\prec n^{-1/2-1/\alpha}.$$

For R_2, applying the Sherman–Morrison formula to $((G^{(i)})^{-1}+\mathbf{y}_i\mathbf{y}_i^*)^{-1}$, we obtain that

$$\text{(S.20)}\quad \frac{1}{n}\left|\operatorname{tr}\Big(\frac{(G^{(i)}-G)\Sigma(I+m_2\Sigma)^{-1}\Sigma}{1+\mathbf{y}_i^*G^{(i)}\mathbf{y}_i}\Big)\right| = \frac{1}{n}\left|\frac{\mathbf{y}_i^*G^{(i)}\Sigma(I+m_2\Sigma)^{-1}\Sigma G\mathbf{y}_i}{1+\mathbf{y}_i^*G^{(i)}\mathbf{y}_i}\right| \prec \frac{\xi_i^2}{n\eta^2},$$

where in the second step we used Lemma S.1.15 and (S.17) and a discussion similar to (S.16). Together with the definition of R_2 in (S.14), by (S.25), we find that

$$\frac{z}{n}|\operatorname{tr}(R_2\Sigma)|\le \frac{1}{n^2}\sum_i \frac{\xi_i^2}{\eta^2}\prec n^{-1-2/\alpha}.$$

As a result, in light of the definition m_1 in (S.24), we have

$$\text{(S.21)}\quad \begin{aligned} m_1 &= \frac{1}{n}\operatorname{tr}(G(z)\Sigma) = -z^{-1}\frac{1}{n}\operatorname{tr}((I+m_2(z)\Sigma)^{-1}\Sigma)+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha})\\ &= -\frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{z(1+m_2\sigma_i)}+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha}). \end{aligned}$$

We first control $m_2(z)-m_{2n}(z)$. Recall $m_{2n}(z)$ in (6.2). Combing (S.12) and (S.21), we have that

$$\text{(S.22)}\quad \begin{aligned} &m_2(z)-m_{2n}(z)\\ &=\frac{1}{n}\sum_{i=1}^n\Big(\frac{\xi_i^2}{-z(1+\xi_i^2m_1)}+\frac{\xi_i^2}{z(1+\xi_i^2m_{1n})}\Big)+\mathrm{O}_{\prec}(n^{-1/2-1/\alpha})\\ &=\frac{1}{n}\sum_{i=1}^n\frac{\xi_i^4(m_1-m_{1n})}{z(1+\xi_i^2m_1)(1+\xi_i^2m_{1n})}+\mathrm{O}_{\prec}(n^{-1/2-1/\alpha})\\ &=\Big(\frac{1}{n}\sum_{i=1}^n\frac{\xi_i^4}{z(1+\xi_i^2m_1)(1+\xi_i^2m_{1n})}\Big)\Big(\frac{1}{n}\sum_{i=1}^n\frac{\sigma_i^2(m_2-m_{2n})}{z(1+\sigma_im_2)(1+\sigma_im_{2n})}\Big)+\mathrm{O}_{\prec}(n^{-1/2-1/\alpha}). \end{aligned}$$

By a discussion similar to (S.16) and (S.17) with high probability, when n is sufficiently large, we have that

$$\begin{aligned} |m_2-m_{2n}| &= \mathrm{O}\Big(\frac{1}{n}\sum_{i=1}^n\frac{\xi_i^4}{z^2}|m_2-m_{2n}|\Big)+\mathrm{O}_{\prec}(n^{-1/2-1/\alpha})\\ &=\mathrm{O}(n^{-2/\alpha}|m_2-m_{2n}|)+\mathrm{O}_{\prec}(n^{-1/2-1/\alpha}), \end{aligned}$$

where in the second step we used (S.25). Then we can conclude that $m_2 - m_{2n} \prec n^{-1/2-1/\alpha}$. By a similar procedure, we also have $m_1 - m_{1n} \prec n^{-1/2-2/\alpha}$.

Armed with the above two results, we proceed to finish the rest of the proof. Recall m_Q in (6.1). Using (S.14) and a discussion similar to (S.21), one can see that

$$m_Q = \frac{1}{p}\mathrm{tr}(G(z)) = -\frac{1}{p}\sum_{i=1}^p \frac{1}{z(1+m_2\sigma_i)} + \mathrm{O}_\prec(n^{-1/2-2/\alpha})$$
$$= -\frac{1}{p}\sum_{i=1}^p \frac{1}{z(1+m_{2n}\sigma_i)} + \frac{1}{p}\sum_{i=1}^p \frac{(m_2-m_{2n})\sigma_i}{z(1+m_2\sigma_i)(1+m_{2n}\sigma_i)} + \mathrm{O}_\prec(n^{-1/2-2/\alpha})$$
$$= m_n + \mathrm{O}_\prec(n^{-1/2-2/\alpha}), \tag{S.23}$$

where in the last step we recall $m_n(z)$ in (6.2). Finally, for the control of the matrix $\mathcal{G}$, for the diagonal entries, by Lemma S.1.13 and a discussion similar to (S.10) and S.11, we have

$$\mathcal{G}_{ii} = -\frac{1}{z(1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i)} = -\frac{1}{z(1+\xi_i^2 n^{-1}\mathrm{tr}G^{(i)}\Sigma + \mathrm{O}_\prec(\frac{\xi_i^2}{\sqrt{n\eta}}))}$$
$$= -\frac{1}{z(1+\xi_i^2 m_1 + \mathrm{O}_\prec(\frac{\xi_i^2}{\sqrt{n\eta}}))} = -\frac{1}{z(1+\xi_i^2 m_{1n})} + \mathrm{O}_\prec(n^{-1/2-1/\alpha}).$$

For off-diagonal entries, together with Lemmas S.1.13 and S.1.15, we have

$$|\mathcal{G}_{ij}| \le |z||\mathcal{G}_{ii}||\mathcal{G}_{ii}^{(i)}||\mathbf{y}_i^* G^{(ij)}\mathbf{y}_j| \prec n^{-1/2-2/\alpha}.$$

This completes the proof when (2.3) holds. For the case (2.4), the main difference is to use the estimates of (S.26) instead of (S.25) whenever it is needed, for example, (S.9). We omit further details. This completes our proof.

□

The second component is to prove Proposition S.2.1 under a priori control of the resolvent which is summarized in the following lemma.

LEMMA S.2.4. *Proposition S.2.1 holds if (S.2) holds uniformly for $z \in \widetilde{\mathbf{D}}_u$.*

PROOF. Note that according to the priori control (S.2), we have that for $1 \leqslant i \neq j \leqslant n$

$$\mathcal{G}_{ii} = \frac{1}{z(1+\xi_i^2 m_{1n}(z)} + \mathrm{O}_\prec(n^{-1/2-1/\alpha}), \quad \mathcal{G}_{ij} = \mathrm{O}_\prec(n^{-1/2-1/\alpha}). \tag{S.24}$$

For the diagonal entries, when $i=1$, using (6.5), we observe that

$$\mathcal{G}_{11} = -\frac{1}{z(1+\xi_1^2 m_{1n}(z))} + \mathrm{O}_\prec(n^{-1/2-1/\alpha})$$

(S.25)

$$= -\frac{1}{z(1+\xi_1^2 m_{1n}(\vartheta_1))} + \frac{z\xi_1^2(m_{1n}(z)-m_{1n}(\vartheta_1))}{(z(1+\xi_1^2 m_{1n}(\vartheta_1)))(z(1+\xi_1^2 m_{1n}(z)))} + \mathrm{O}_\prec(n^{-1/2-1/\alpha})$$
$$= \frac{1}{zd_1 m_{1n}(\vartheta_1)} - \frac{z\xi_1^2(m_{1n}(z)-m_{1n}(\vartheta_1))}{zd_1 m_{1n}(\vartheta_1)}\big(\mathcal{G}_{11} + \mathrm{O}_\prec(n^{-1/2-2/\alpha})\big) + \mathrm{O}_\prec(n^{-1/2-1/\alpha})$$
$$= \frac{1}{zd_1 m_{1n}(\vartheta_1)} - \frac{z\xi_1^2}{zd_1 m_{1n}(\vartheta_1)}(\mathcal{G}_{11} + \mathrm{O}_\prec(n^{-1/2-1/\alpha})) \times \mathrm{O}_\prec(n^{-1/\alpha}) + \mathrm{O}_\prec(n^{-1/2-1/\alpha}),$$

where in the fourth step we used Lemma S.1.2. By Lemma S.1.2, (S.8) and (S.25), this yields that for some constant $C>0$

$$|\mathcal{G}_{11}| = \frac{1}{|zd_1 m_{1n}(\vartheta_1)|} + \mathrm{O}_{\prec}(n^{-1/2-1/\alpha}) = \frac{C}{d_1} + \mathrm{O}_{\prec}(n^{-1/2-1/\alpha}). \tag{S.26}$$

Similarly, when $2 \leqslant i \leqslant n$, by (S.25) and the definition of d_1, using Lemma S.1.2, we see that

$$\mathcal{G}_{ii} = \mathrm{O}_{\prec}(n^{-1/\alpha}). \tag{S.27}$$

We also provide some basic controls for the matrix $\mathcal{G}^{(i)}$ for all $1 \leqslant i \leqslant n$. By definition and an elementary calculation, it is not hard to see that

$$\mathcal{G}_{ii}^{(i)} = -z^{-1};\ \mathcal{G}_{ik}^{(i)} = 0,\ 1 \leqslant k \neq i \leqslant n. \tag{S.28}$$

Moreover, using (S.24), (S.27) and the third identity of Lemma S.1.13, we find that for $1 \leqslant i \leqslant n$,

$$\mathcal{G}_{kk}^{(i)} = \mathrm{O}_{\prec}(n^{-1/\alpha}),\ k \neq i;\ \mathcal{G}_{kl}^{(i)} = \mathrm{O}_{\prec}(n^{-1/2-1/\alpha}),\ k,l \neq i; \tag{S.29}$$

With the above preparation, we now proceed to the control of Z_i in (S.9), unlike in (S.10), since Q and $\mathcal{Q}$ have the same non-zero eigenvalues, we control it as follows using the above bounds

$$\begin{aligned}
Z_i &\prec \frac{\xi_i^2}{n}\|G^{(i)}\Sigma\|_F \prec \frac{\xi_i^2}{n}\|G^{(i)}\|_F = \frac{\xi_i^2}{n}(\mathrm{tr}((G^{(i)})^2))^{1/2} \leq \frac{\xi_i^2}{n}\mathrm{tr}((\mathcal{G}^{(i)})^2)^{1/2} + \frac{\xi_i^2}{n}\frac{\sqrt{|n-p|}}{|z|} \\
&\asymp \frac{\xi_i^2}{n}\|\mathcal{G}^{(i)}\|_F + \frac{\xi_i^2}{n}\frac{n^{1/2}}{n^{1/\alpha}} = \frac{\xi_i^2}{n}\Big((\mathcal{G}_{ii}^{(i)})^2 + \sum_{j\neq i}(\mathcal{G}_{jj}^{(i)})^2 + \sum_{j\neq k\neq i}(\mathcal{G}_{jk}^{(i)})^2\Big)^{1/2} + \frac{\xi_i^2}{n^{1/2+1/\alpha}} \\
&\prec \frac{\xi_1^2}{n}\Big(|z|^{-2} + nn^{-2/\alpha} + n^2 n^{-1-2/\alpha}\Big)^{1/2} + \frac{\xi_i^2}{n^{1/2+1/\alpha}} \prec \frac{\xi_i^2}{n^{1/2+1/\alpha}},
\end{aligned} \tag{S.30}$$

where in the third and fourth steps we used (S.28) and (S.29).

Besides, unlike in (S.11), by a discussion similar to (S.30), we now have from Lemma S.1.15 that

$$\begin{aligned}
T_i := \frac{1}{n}\mathrm{tr}G^{(i)}\Sigma - m_1(z) = \frac{1}{n}\mathbf{y}_i^* G\Sigma G^{(i)}\mathbf{y}_i &= \frac{1}{n}\frac{\mathbf{y}_i^* G^{(i)}\Sigma G^{(i)}\mathbf{y}_i}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i} \\
&\prec \frac{\xi_i^2}{n}\frac{n^{-1}\|G^{(i)}\|_F^2}{|1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i|} \\
&\prec \frac{\xi_i^2}{n^2}|z||\mathcal{G}_{ii}|\left(\|\mathcal{G}^{(i)}\|_F^2 + \frac{n}{n^{2/\alpha}}\right), \\
&\prec \frac{\xi_i^2}{n^{1+2/\alpha}}.
\end{aligned} \tag{S.31}$$

where in the second step we used the relation (S.13), in the third step we used Lemma S.1.13 and in the last two steps we used a discussion similar to (S.30).

With the above control, we now use an idea similar to the proof of Lemma S.2.3 to conclude the proof. The key ingredient is to explore the relation of m_1 and m_2. We start with m_2. Using the above notations and Lemma S.1.13, we find that

$$\frac{1}{-z(1+\xi_i^2 m_1(z))} = \frac{1}{\mathcal{G}_{ii}^{-1} + z(Z_i+T_i)}. \tag{S.32}$$

Consequently, by (S.26), (S.27), (S.30) and (S.31), we see that

$$\frac{1}{-z(1+\xi_i^2 m_1(z))} \prec n^{-1/\alpha}. \tag{S.33}$$

Then using the decomposition as in (S.9), we have that

$$\begin{aligned} m_2 &= \frac{1}{n}\frac{\xi_1^2}{-z(1+\xi_1^2 n^{-1}\operatorname{tr} G^{(1)}\Sigma + Z_1)} + \frac{1}{n}\sum_{i=2}^n \frac{\xi_i^2}{-z(1+\xi_i^2 n^{-1}\operatorname{tr} G^{(i)}\Sigma + Z_i)} \\ &= \frac{1}{n}\frac{\xi_1^2}{-z(1+\xi_1^2 m_1(z)+\xi_1^2 n^{-1-2/\alpha} + Z_1)} + \frac{1}{n}\sum_{i=2}^n \frac{\xi_i^2}{-z(1+\xi_i^2 m_1(z)+\xi_i^2 n^{-1-2/\alpha} + Z_i)} \qquad \text{(S.34)} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^2}{-z(1+\xi_i^2 m_1(z))} + \mathrm{O}_{\prec}\Big(n^{-3/2} + \frac{1}{n}\sum_{i=2}^n \frac{\xi_i^4}{|z| n^{1/2+1/\alpha}}\Big) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^2}{-z(1+\xi_i^2 m_1(z))} + \mathrm{O}_{\prec}(n^{-1/2-1/\alpha}), \end{aligned}$$

where in the second step we used (S.31), in the third step we used a discussion similar to (S.25) and in the last step we used (S.25). Then we study m_1 using the arguments as between (S.14) and (S.21). We provide the key ingredients as follows. First, for $\mathtt{R}_{11}$ in (S.15), note that by definition of $m_2^{(i)}$ and (S.28), we have that

$$\begin{aligned} \big|m_2^{(i)}\big| &\le \frac{1}{n}\Big(\sum_{j\neq i} \xi_j^2 |\mathcal{G}_{jj}^{(i)}| + |z|^{-1}\Big) \\ &\le \frac{1}{n}\Big[\sum_{j\neq 1}\xi_j^2\Big(|\mathcal{G}_{jj}| + \frac{|\mathcal{G}_{j1}||\mathcal{G}_{1j}|}{|\mathcal{G}_{11}|}\Big) + |z|^{-1}\Big] \\ &\prec n^{-1/\alpha}, \end{aligned}$$

where in the second step we used Lemma S.1.13 and in the last step we used (S.24), (S.27) and (S.25). Moreover, by Lemma S.1.13 and (S.27), we find that $(1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i)^{-1} \prec 1$. Therefore, we conclude that for all $1 \leqslant i \leqslant n$,

$$\begin{aligned} &\operatorname{tr}\Big(\frac{G^{(i)}(\mathbf{y}_i\mathbf{y}_i^* - n^{-1}\xi_i^2\Sigma)}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}(I+m_2^{(i)}\Sigma)^{-1}\Sigma\Big) \asymp \operatorname{tr}\Big(\xi_i^2 G^{(i)}(\mathbf{u}_i\mathbf{u}_i^* - n^{-1}I)(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\Big) \\ &= \xi_i^2\Big(\mathbf{u}_i^* G^{(i)}(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\mathbf{u}_i - n^{-1}\operatorname{tr}\Big(G^{(i)}(I+m_2^{(i)}\Sigma)^{-1}\Sigma^2\Big)\Big) \\ &\prec \frac{\xi_i^2}{n}\|G^{(i)}\|_F \prec \frac{\xi_i^2}{n^{1/2+1/\alpha}}, \end{aligned}$$

where in the last step we used a discussion similar to (S.30). Together with (S.25), we can conclude that $\mathtt{R}_{11} \prec n^{-1/2-1/\alpha}$. For $\mathtt{R}_{12}$, using (S.19), (S.24) and (S.27)

$$m_2 - m_2^{(i)} \prec \frac{1}{n}\sum_{j\neq i}\frac{\xi_i^2 n^{-1-2/\alpha}}{n^{-1/\alpha}} + \frac{1}{n} \prec n^{-1}.$$

Then by an argument similar to (S.20), we can conclude that $\mathtt{R}_{12} \prec n^{-1-1/\alpha}$. Similarly, for R_2, we have that

$$\frac{1}{n}\left|\operatorname{tr}\left(\frac{(G^{(i)}-G)\Sigma(I+m_2\Sigma)^{-1}\Sigma}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}\right)\right| = \frac{1}{n}\left|\frac{\mathbf{y}_i^* G^{(i)}\Sigma(I+m_2\Sigma)^{-1}\Sigma G\mathbf{y}_i}{1+\mathbf{y}_i^* G^{(i)}\mathbf{y}_i}\right|$$
$$\asymp \frac{1}{n}\left|\mathbf{y}_i^* G^{(i)}\Sigma(I+m_2\Sigma)^{-1}\Sigma G^{(i)}\mathbf{y}_i\right| \prec \frac{\xi_i^2}{n^2}\|G^{(i)}\|_F\|G\|_F \prec \frac{\xi_i^2}{n^{1+2/\alpha}}.$$

Consequently, we have that

$$\frac{z}{n}\left|\operatorname{tr}(R_2\Sigma)\right| \prec \frac{1}{n}\sum_i \frac{\xi_i^2}{n^{1+1/\alpha}} \prec n^{-1-1/\alpha}.$$

Combining all the above arguments, we find that (S.21) still holds true. Armed with all the above controls, using an argument similar to the discussions between (S.22) and (S.23), we can conclude the proof. □

Combining the above two lemmas, we now proceed to the proof of Proposition S.2.1. We will use a continuity argument as in [26, Lemma A.12] or [13, Section 4.1]. In fact, our discussion is easier since the real part in the spectral domain $\widetilde{\mathbf{D}}_u$ is divergent so that the rate is independent of η. Due to similarity, we focus on explaining the key ingredients.

Proof of Proposition S.2.1. For each $z = E + \mathrm{i}\eta \in \widetilde{\mathbf{D}}_u$, we fix the real part and construct a sequence $\{\eta_j\}$ by setting $\eta_j = \mathtt{C}\vartheta_1 - jn^{-3}$. Then it is clear that η falls in an interval $[\eta_{j-1}, \eta_j]$ for some $0 \le j \le Cn^{1/\alpha+3}, C > 0$ is some constant.

In Lemma S.2.3, we have proved that the results hold for η_0. Now we assume (S.2) holds for some η_j. Then according to Lemma S.2.4, we have that

$$|m_1(z_j) - m_{1n}(z_j)| + |m_Q(z_j) - m_n(z_j)| \prec n^{-1/2-2/\alpha}, \; |m_2(z_j) - m_{2n}(z_j)| \prec n^{-1/2-1/\alpha}.$$

For any η' lying in the interval $[\eta_{k-1}, \eta_k]$, denote $z' = E + \mathrm{i}\eta'$ and $z_j = E + \mathrm{i}\eta_j$. According to the first resolvent identity, we have that

$$\|\mathcal{G}(z') - \mathcal{G}(z_j)\| \leqslant n^{-3}\|\mathcal{G}(z')\|\|\mathcal{G}(z_j)\| \prec n^{-11/6-1/\alpha}, \tag{S.35}$$

where in the second step we used the basic bound $\|\mathcal{G}(z')\| \leqslant n^{2/3}$, (S.24), (S.27) and Gershgorin circle theorem.

On the one hand, according to the definitions in (S.24), using the first resolvent identity, we have that

$$m_1(z') - m_1(z_j) = \frac{1}{n}\operatorname{tr}[(G(z') - G(z_j))\Sigma] = \frac{1}{n^4}\operatorname{tr}(G(z')G(z_j)\Sigma) \prec \frac{1}{n^4\eta_j}\|G(z_j)\|_F \prec n^{-17/6-1/\alpha},$$

where in the last step we used a discussion similar to (S.30). Similarly, combining (S.35) and (S.25), we have that

$$m_2(z') - m_2(z_j) = \frac{1}{n}\sum_{i=1}^n \xi_i^2(\mathcal{G}_{ii}(z') - \mathcal{G}_{ii}(z_j)) \prec n^{-11/6-1/\alpha},$$

and by a discussion similar to (S.30)

$$|m_Q(z') - m_Q(z_j)| = \frac{1}{p}\left|\operatorname{tr}\left(G(z') - G(z_j)\right)\right| \leqslant n^{-4}\|G(z')\|_F\|G(z_j)\|_F$$
$$\prec n^{-4}(n^{1-2/\alpha})^{1/2}n^{1/2+2/3} = n^{-7/3-1/\alpha}.$$

On the other hand, using the definitions in (6.2), we decompose that

$$\begin{aligned} m_{1n}(z') - m_{1n}(z_j) &= \frac{1}{n}\sum_i \Big(\frac{\sigma_i}{-z'(1+\sigma_i m_{2n}(z'))} - \frac{\sigma_i}{-z'(1+\sigma_i m_{2n}(z_j))} \Big) \\ &\quad + \frac{1}{n}\sum_i \Big(\frac{\sigma_i}{-z'(1+\sigma_i m_{2n}(z_j))} - \frac{\sigma_i}{-z_j(1+\sigma_i m_{2n}(z_j))} \Big) \\ &:= \mathcal{M}_{11} + \mathcal{M}_{12}. \end{aligned}$$

For $\mathcal{M}_{11}$, according to Lemma S.1.2 and (6.2), we readily obtain that

$$\begin{aligned} \mathcal{M}_{11} &= \frac{1}{n}\sum_i \frac{\sigma_i^2}{-z'(1+\sigma_i m_{2n}(z'))(1+\sigma_i m_{2n}(z_j))} (m_{2n}(z_j) - m_{2n}(z')) \\ &= \mathrm{O}(|z'|^{-1}) \times \frac{1}{n}\sum_i \Big(\frac{\xi_i^2}{-z_j(1+\xi_i^2 m_{1n}(z_j))} - \frac{\xi_i^2}{-z_j(1+\xi_i^2 m_{1n}(z'))} \Big) \\ &\quad + \mathrm{O}(|z'|^{-1}) \times \frac{1}{n}\sum_i \Big(\frac{\xi_i^2}{-z_j(1+\xi_i^2 m_{1n}(z'))} - \frac{\xi_i^2}{-z'(1+\xi_i^2 m_{1n}(z'))} \Big) \\ &= \mathrm{O}(|z'|^{-1}) \times (\mathtt{M}_{11,1} + \mathtt{M}_{11,2}). \end{aligned}$$

For $\mathtt{M}_{11,1}$, by a discussion similar to (S.26) and (S.27), we find that

$$\begin{aligned} \mathtt{M}_{11,1} &= \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^4}{-z_j(1+\xi_i^2 m_{1n}(z'))(1+\xi_i^2 m_{1n}(z_j))} (m_{1n}(z') - m_{1n}(z_j)) \\ &\prec \Big(\frac{\xi_1^4}{-nz_j d_1^2 |m_{1n}(z_j)||m_{1n}(z')|} + \frac{1}{n}\sum_{i\ge 2} \frac{\xi_i^4}{|z'(1+\xi_1^2 m_{1n}(z'))(1+\xi_i^2 m_{1n}(z_j))|} \Big) \times |m_{1n}(z_j) - m_{1n}(z')| \\ &\prec \mathrm{o}(1) \times |m_{1n}(z_j) - m_{1n}(z')|. \end{aligned}$$

Similarly, for $\mathtt{M}_{11,2}$, we have that

$$\mathtt{M}_{11,2} = \frac{1}{n}\sum_i \frac{\xi_i^2 (z_j - z')}{z' z_j (1+\xi_i^2 m_{1n}(z'))} \prec \frac{z' - z_j}{z' z_j} \prec n^{-3-2/\alpha}.$$

Analogously, we can prove that $\mathcal{M}_{12} \prec n^{-3-4/\alpha}$. Therefore, combining the above bounds with (S.25), we see that

$$|m_{1n}(z') - m_{1n}(z_j)| \prec n^{-3-2/\alpha}.$$

By similar procedures and arguments, we can also prove that

$$|m_{2n}(z') - m_{2n}(z_j)| \prec n^{-3-2/\alpha}, \ |m_n(z') - m_n(z_j)| \prec n^{-3-2/\alpha}.$$

and

$$\big\| (z')^{-1}(I + m_{1n}(z')D^2)^{-1} - (z_j)^{-1}(I + m_{1n}(z_j)D^2)^{-1} \big\| \prec n^{-3-2/\alpha}.$$

Therefore, combining all the above bounds with triangle inequality, we see that the results of part 1 of Theorem S.1.8 hold for z'. Using an induction procedure and a standard lattice argument (for example, see [13, 26]), we find that the results hold for all $z \in \widetilde{\mathbf{D}}_u$ and conclude the proof of Proposition S.2.1. □

S.2.1.2. Proof of Theorem S.1.8

Once Proposition S.2.1 is proved, we can roughly locate the edge eigenvalues of Q as in Lemma S.2.2 so that we can expand the spectral domain from $\widetilde{\mathbf{D}}_u$ to $\mathbf{D}_u$ for $Q^{(1)}$ and conclude the proof of Theorem S.1.8.

Recall the definitions of μ_2 and $\lambda_1^{(1)}$ around (S.1) and in Figure S.3. By Lemma S.2.2 and an analogous argument, as well as Weyl's inequality, we find that conditional on the event Ω, with high probability,

$$\vartheta_1 > \lambda_1 > \vartheta_2 > \lambda_1^{(1)}. \tag{S.36}$$

By (S.8) and a similar argument, we see that $\vartheta_k \asymp \xi_k^2, k = 1, 2$. Together with (S.25), we have that $\vartheta_1 - \vartheta_2 \geq C_1 n^{1/\alpha} \log^{-1} n$ for some constant $C_1 > 0$ on the event Ω. This implies for some constant $C > 0$, for $z \in \mathbf{D}_u$,

$$|\lambda_1^{(1)} - z| \geq C n^{1/\alpha} \log^{-1} n, \tag{S.37}$$

Now we proceed to the proof of Theorem S.1.8. Recall (S.7) and (S.8).

Proof of Theorem S.1.8. Observe by (S.37) that it holds uniformly for $z \in \mathbf{D}_u$ and $\mathcal{T} \subset \{2, \ldots, n\}$, for some constant $C_1 > 0$

$$\|G^{(1\mathcal{T})}\| \leqslant C_1 n^{-1/\alpha} \log n. \tag{S.38}$$

By the definition of $m_2^{(1)}$ and a decomposition similar to (S.9), we have that

$$m_2^{(1)} = \frac{1}{n} \sum_{i=2}^{n} \frac{\xi_i^2}{-z - z\mathbf{y}_i^* G^{(1i)} \mathbf{y}_i} = \frac{1}{n} \sum_{i=2}^{n} \frac{\xi_i^2}{-z(1 + \xi_i^2 n^{-1} \operatorname{tr} G^{(1i)} \Sigma + Z_i^{(1)})},$$
$$Z_i^{(1)} = \mathbf{y}_i^* G^{(1i)} \mathbf{y}_i - \xi_i^2 n^{-1} \operatorname{tr} G^{(1i)} \Sigma.$$

By arguments similar to (S.10) and (S.11) but with (S.38), we obtain that

$$Z_i^{(1)} \prec \frac{\xi_i^2}{n} \|G^{(1i)} \Sigma\|_F \leq \frac{\xi_i^2}{n} \|G^{(1i)}\| \|\Sigma\|_F \prec \frac{\xi_i^2}{n^{1/2}} n^{-1/\alpha},$$
$$\frac{1}{n} \operatorname{tr}(G^{(1i)} \Sigma) - m_1^{(1)}(z) = \frac{1}{n} \mathbf{y}_i^* G^{(1)} \Sigma G^{(1i)} \mathbf{y}_i \prec \frac{\xi_i^2}{n} n^{-2/\alpha}.$$

In addition, using (S.38) and a discussion similar to (S.32)–(S.34), we readily see that

$$m_2^{(1)} = \frac{1}{n} \sum_{i=2}^{n} \frac{\xi_i^2}{-z(1 + \xi_i^2 m_1^{(1)})} + \mathrm{O}_{\prec}(n^{-1/2-1/\alpha}).$$

Using the decomposition

$$Q^{(1)} - zI = \sum_{i=2}^{n} \mathbf{y}_i \mathbf{y}_i^* + z m_2^{(1)}(z) \Sigma - z(I + m_2^{(1)}(z) \Sigma),$$

by arguments similar to (S.14)–(S.21) with $\|\mathcal{G}^{(1\mathcal{T})}\| = \|G^{(1\mathcal{T})}\| \prec n^{-1/\alpha}$, we conclude that

$$m_1^{(1)} = -z^{-1} \frac{1}{n} \operatorname{tr}((I + m_2^{(1)}(z)\Sigma)^{-1} \Sigma) + \mathrm{O}_{\prec}(n^{-1/2-2/\alpha})$$
$$= -\frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i}{z(1 + m_2^{(1)} \sigma_i)} + \mathrm{O}_{\prec}(n^{-1/2-2/\alpha}).$$

Combining with the definitions in (6.2), we see that

$$m_1^{(1)}(z)-m_{1n}^{(1)}(z)=-\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i}{z(1+\sigma_i m_2^{(1)}(z))}+\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i}{z(1+\sigma_i m_{2n}^{(1)}(z))}+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha})$$
$$=\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i^2(m_2^{(1)}(z)-m_{2n}^{(1)}(z))}{z(1+\sigma_i m_{2n}^{(1)}(z))(1+\sigma_i m_2^{(1)}(z))}+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha})$$
$$=\Big(\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i^2}{z(1+\sigma_i m_{2n}^{(1)}(z))(1+\sigma_i m_2^{(1)}(z))}\Big)\Big(\frac{1}{n}\sum_{i=2}^n\frac{\xi_i^4(m_1^{(1)}(z)-m_{1n}^{(1)}(z))}{z(1+\xi_i^2 m_1^{(1)}(z))(1+\xi_i^2 m_{1n}^{(1)}(z))}\Big)+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha})$$
$$=\mathrm{o}(1)(m_1^{(1)}(z)-m_{1n}^{(1)}(z))+\mathrm{O}_{\prec}(n^{-1/2-2/\alpha}),$$

where in the third step we used a discussion similar to (S.33) and (S.25). This completes our proof. □

S.2.2. Bounded support setting: proof of Theorem S.1.9

In this section, we will prove Theorem S.1.9. In Section S.2.2.1, we study $m_{1n,c}(z)-m_{1n}(z)$ and $m_{n,c}-m_n$, which is a counterpart of Lemma 4.4 of [52]. Then in Section S.2.2.2, we study $m_{1n}(z)-m_1(z)$ and m_Q-m_n, which is a counterpart of Proposition 5.1 of [52].

S.2.2.1. Control of $m_{1n,c}(z)-m_{1n}(z)$ and $m_{n,c}(z)-m_n(z)$

Due to similarity, we focus on $|m_{1n,c}-m_{1n}|$ and briefly discuss $|m_{n,c}-m_n|$ in the end. The proof ideas follow Lemma 4.5 of [56] or Lemma 4.4 of [52]. We focus on explaining the parts deviates the most.

PROOF. According to the definitions of $m_{1n,c}$ and m_{1n} in (6.3) and (6.2), we observe that

(S.39)

$$|m_{1n,c}(z)-m_{1n}(z)|$$
$$\leqslant\Big|\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i}{-z+\sigma_i\int\frac{s}{1+sm_{1n,c}(z)}\mathrm{d}F(s)}-\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i}{-z+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)}}\Big|$$

(S.40)

$$+|m_{1n,c}(z)-m_{1n}(z)|\Big|\frac{1}{n}\sum_{i=1}^p\frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n\frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(z))(1+\xi_j^2 m_{1n,c}(z))}}{(-z+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})(-z+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)})}\Big|$$
$$:=\mathsf{P}_1+\mathsf{P}_2.$$

On the one hand, for P_1, we have that

$$\mathsf{P}_1=\frac{1}{n}\sum_{i=1}^p\frac{\sigma_i^2|n^{-1}\sum_j\frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)}-\int\frac{s}{1+sm_{1n,c}(z)}\mathrm{d}F(s)|}{|(-z+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)})(-z+\sigma_i\int\frac{s}{1+sm_{1n,c}(z)}\mathrm{d}F(s))|}.$$

Since $z\in\mathbf{D}_b'\subset\mathbf{D}_b$, according to Assumption S.1.1 and the continuity of $m_{1n,c}$, we conclude that $|-z+\sigma_i\int\frac{s}{1+sm_{1n,c}(z)}\mathrm{d}F(s))|\geqslant c$ for some constant $c>0$. Together with (S.27), we show that $|-z+\frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)}| \geqslant c'$ for some $c'>0$ when n is sufficiently large. Using (S.27) again, we conclude that on Ω, for some small constant $\epsilon>0$ and some constant $C>0$

$$\mathsf{P}_1 \leqslant C n^{-1/2+\epsilon}. \tag{S.41}$$

On the other hand, for P_2, for notional convenience, we further write it as $\mathsf{P}_2 = |m_{1n,c}(z)-m_{1n}(z)| \times |\mathsf{T}|$. For T, by Cauchy-Schwarz inequality, we have that

$$|\mathsf{T}| \leqslant \mathsf{E}_1 \mathsf{E}_2, \tag{S.42}$$

where $\mathsf{E}_k, k=1,2$, are defined as

$$\begin{aligned} \mathsf{E}_1 &:= \Big(\frac{1}{n}\sum_{i=1}^p \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n \frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(z))^2}}{|-z+\frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2} \Big)^{1/2}, \\ \mathsf{E}_2 &:= \Big(\frac{1}{n}\sum_{i=1}^p \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n \frac{\xi_j^4}{(1+\xi_j^2 m_{1n,c}(z))^2}}{|-z+\frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)}|^2} \Big)^{1/2}. \end{aligned} \tag{S.43}$$

Together with the identity (S.21) below and the fact $m_{1n}(z) \asymp 1$ for $z \in \mathbf{D}_b'$, we find that $\mathsf{E}_1 \leqslant 1$. For the term E_2, we first consider a closely related quantity $\mathsf{W}(z)$ defined as

$$\mathsf{W}(z) := \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i^2 \int \frac{s^2}{|1+s m_{1n,c}(z)|^2} \mathrm{d}F(s)}{|-z+\sigma_i \int \frac{s^2}{(1+s m_{1n,c}(z))^2}\mathrm{d}F(s)|^2} = 1-\eta \frac{|m_{1n,c}(z)|^2}{\operatorname{Im} m_{1n,c}(z)}.$$

By assumption that $\phi^{-1} > \mathsf{s}_3$, (3.6) that $m_{1n,c}(L_+) = -l^{-1}$ and recall the notations in (3.5), we see that

$$\mathsf{W}(L_+) < 1. \tag{S.44}$$

Armed with (S.44), using (S.27) and Assumption S.1.1, we can apply an argument similar to Lemma A.6 of [56] or Lemma A.7 of [52] to conclude that when n is sufficiently large, for $z \in \mathbf{D}_b$,

$$\mathsf{E}_2^2 = \mathsf{W}(L_+) + \mathrm{o}(1) < \mathfrak{c}', \tag{S.45}$$

for some constant $0 < \mathfrak{c}' < 1$.

Consequently, we find that when n is sufficiently large, $\mathsf{E}_2 < 1$. Together with (S.42), we can conclude that $|\mathsf{T}| < 1$. This yields that

$$\mathsf{P}_2 = \mathfrak{c}|m_{1n,c} - m_{1n}|,$$

for some constant $0 < \mathfrak{c} < 1$.

Inserting the above control back into (S.39), using (S.41), we can conclude our proof that

$$m_{1n,c} = m_{1n}(z) + \mathrm{O}(n^{-1/2+\epsilon}). \tag{S.46}$$

The proof of $m_{n,c} - m_n$ follows from an argument similar to (S.39) using (6.2) and (6.3) that

$$m_n(z) = \frac{1}{p}\sum_{i=1}^p \frac{1}{-z+\sigma_i n^{-1}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}}}, \quad m_{n,c}(z) = \frac{1}{p}\sum_{i=1}^p \frac{1}{-z+\sigma_i \int_0^l \frac{s}{1+s m_{1n,c}(z)}\mathrm{d}F(s)},$$

the results of $m_{1n,c} - m_{1n}$ and (S.27). We omit the details. □

S.2.2.2. Control of $m_{1n}(z) - m_1(z)$ and $m_n(z) - m_Q(z)$

Due to similarity, we focus on $|m_{1n,c} - m_{1n}|$ and will briefly discuss how to study $|m_Q - m_n|$ from line to line. The proof ideas follow Proposition 5.1 of [56] or Proposition 5.1 of [52]. We focus on explaining the parts deviates the most. The proof relies on the following two lemmas.

LEMMA S.2.5. *Conditional on the event Ω in Theorem S.1.9, for all $z = E + \mathrm{i}\eta \in \mathbf{D}_b'$ with $n^{-1/2+\epsilon_d} \le \eta \le n^{-1/(d+1)+\epsilon_d}$, we have*

$$|m_{1n}(z) - m_1(z)| \prec \frac{1}{n\eta_0}, \ |m_n(z) - m_Q(z)| \prec \frac{1}{n\eta_0}.$$

LEMMA S.2.6. *Assuming that $|m_{1n}(z) - m_1(z)| \prec n^{\epsilon_d}(n\eta_0)^{-1}$, then conditional on the event Ω, we have that for all $z \in \mathbf{D}_b'$*

$$|m_{1n}(z) - m_1(z)| \prec \frac{1}{n\eta_0}, \ |m_n(z) - m_Q(z)| \prec \frac{1}{n\eta_0}.$$

Armed with the above two lemmas, we now proceed to the control of $m_{1n}(z) - m_1(z)$.

Proof: control of $m_{1n}(z) - m_1(z)$ and $m_n(z) - m_Q(z)$. Due to similarity, we only prove $m_{1n}(z) - m_1(z)$. We prove this by mathematical induction as that of Proposition 5.1 of [56]. Fix E such that $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$, we consider a sequence (η_j) defined by $\eta_j = \eta_0 + jn^{-2}$. Let K be the smallest positive integer such that $\eta_K \ge n^{-1/2+\epsilon_d}$. Note that for $j = K$, by Lemma S.2.5, we have that

$$|m_{1n}(z_j) - m_1(z_j)| \prec \frac{1}{n\eta_0}.$$

Then for any $z = E + \mathrm{i}\eta$ with $\eta_{j-1} \le \eta \le \eta_j$, we have that for some constant $C > 0$

$$|m_1(z_j) - m(z)| = \frac{1}{n}\operatorname{tr}\left[(G(z_j) - G(z))\Sigma\right] = \frac{|z_j - z|}{n}\operatorname{tr}(G(z_j)G(z)\Sigma) \le C\frac{|z_j - z|}{\eta_{j-1}^2} \le C\frac{n^{2\epsilon_d}}{n},$$

where we used the first resolvent identity and the trivial bound $|G(z)| \leqslant \eta^{-1}$, and similarly

$$|m_{1n}(z_j) - m_{1n}(z)| = \left|\int \left[\frac{1}{x - z_j} - \frac{1}{x - z}\right]\rho(x)\mathrm{d}x\right| \le \frac{|z_j - z|}{\eta_{j-1}^2} \le \frac{n^{2\epsilon_d}}{n}.$$

Thus we find that if $|m_{1n}(z_j) - m_1(z_j)| \prec \frac{1}{n\eta_0}$, then by Lemma S.2.6, for some constant $C' > 0$

$$|m_{1n}(z) - m_1(z)| \le |m_{1n}(z_j) - m_1(z_j)| + \frac{C'n^{2\epsilon_d}}{n} \prec \frac{n^{\epsilon_d}}{n\eta_0}. \tag{S.47}$$

This gives the result that $|m_{1n}(z) - m_1(z)| \prec (n\eta_0)^{-1}$ for $z = E + \mathrm{i}\eta$ with $\eta_{j-1} \le \eta \le \eta_j$. The proof for each z can be completed by an induction on j. Finally, using an induction procedure and a standard lattice argument (for example, see [13, 26]), we find that the results hold for all $z \in \mathbf{D}_b'$. More specifically, we construct a lattice $\mathcal{L}$ from $z' = E' + \mathrm{i}\eta_0 \in \mathbf{D}_b'$ with $|z - z'| \le n^{-3}$. It is obvious that the bound holds uniformly on $\mathcal{L}$. For any $z = E + \mathrm{i}\eta_0 \notin \mathcal{L}$, we find a $z' \in \mathcal{L}$ and then $|z - z'| \le n^{-3}$. Moreover, using resolvent identity, we can conclude that $|m_1(z) - m_1(z')| \le \eta_0^{-2}|z - z'| \ll (n\eta_0)^{-1}$. Therefore, we conclude the proof.

□

In what follows, we prove lemmas S.2.5 and S.2.6. The proofs are similar to those of Lemmas 5.6 and 5.7 of [56], except that we will need a weak local law as follows.

PROPOSITION S.2.7 (Weak averaged local law). *Suppose the assumptions of Theorem S.1.9 hold. We have that for $z \in \mathbf{D}_b'$*

$$|m_Q(z) - m_n(z)| + |m_1(z) - m_{1n}(z)| + |m_2(z) - m_{2n}(z)| = \mathrm{O}_{\prec}\left((n\eta)^{-1/4}\right).$$

PROOF. The proof of Proposition S.2.7 is relatively standard in the random matrix literature, for example, see Section 4.1 of [13] or Section 3.6 of [35] or Appendix A.2 of [26] or Section 5.2 of [71]. Due to similarity, as in Lemma 5.12 of [71], we only provide the key ingredients. Define the z-dependent parameter

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_1(z)}{n\eta}} + \frac{1}{n\eta}. \tag{S.48}$$

Recall (S.9). By Lemma S.1.15 and (S.28), we find that

$$Z_i \prec \frac{\xi_i^2}{n}\|G^{(i)}\Sigma^{1/2}\|_F \leqslant l\sqrt{\frac{\operatorname{Im} m_1^{(i)}(z)}{n\eta}} \asymp \Psi, \tag{S.49}$$

where in the last step we used (S.29). Together with (S.29) and the first equation of (S.9), we conclude that

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i^2}{-z(1+\xi_i^2 m_1(z) + \mathrm{O}_{\prec}(\Psi))}. \tag{S.50}$$

For $m_1(z)$, recall (S.14). According to the definition of $m_1(z)$ in (S.24), we have that

$$m_1(z) = \frac{1}{n}\operatorname{tr}(G(z)\Sigma) = -\frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{z(1+m_2(z)\sigma_i)} + \frac{1}{n}\operatorname{tr}(R_1\Sigma) + \frac{1}{n}\operatorname{tr}(R_2\Sigma). \tag{S.51}$$

Similarly, for $m_Q(z)$ in (6.1), we have that

$$m_Q(z) = \frac{1}{p}\operatorname{tr}(G(z)) = -\frac{1}{p}\sum_{i=1}^{p} \frac{1}{z(1+m_2(z)\sigma_i)} + \frac{1}{p}\operatorname{tr}(R_1) + \frac{1}{p}\operatorname{tr}(R_2). \tag{S.52}$$

On the one hand, when $\eta \asymp 1$, by a discussion similar to (5.45) of [71] , we find that $\|(I + m_2^{(i)}\Sigma)^{-1}\| < \infty$. Then using (S.49), by a discussion similar to the equations between (S.15) and (S.21), we find that

$$m_1(z) = -\frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{z(1+m_2(z)\sigma_i)} + \mathrm{O}_{\prec}\Big(\frac{1}{n}\sum_i \frac{\xi_i^2\Psi}{z(1+\xi_i^2 m_1(z) + \mathrm{O}_{\prec}(\Psi))}\Big). \tag{S.53}$$

Similarly, we have

$$m_Q(z) = -\frac{1}{p}\sum_{i=1}^{p} \frac{1}{z(1+m_2(z)\sigma_i)} + \mathrm{O}_{\prec}\Big(\frac{1}{p}\sum_i \frac{\xi_i^2\Psi}{z(1+\xi_i^2 m_1(z) + \mathrm{O}_{\prec}(\Psi))}\Big).$$

On the other hand, denote $\Xi := \{|\mathcal{G}_{ij}(z) + \delta_{ij}(z(1+m_{1n}(z)\xi_i^2))^{-1}| + |m_2(z) - m_{2n}(z)| \leqslant (\log n)^{-1}\}$. When restricted on Ξ, by Assumption S.1.1, we also have that $\|(I + m_2^{(i)}\Sigma)^{-1}\| < \infty$. By an analogous argument, we find that (S.53) also holds true. By an argument similar to (S.29) using Lemma S.1.13, we have that for $i \neq j$, $\mathbf{1}(\eta \geqslant 1)\mathcal{G}_{ij} \prec \Psi$, $\mathbf{1}(\Xi)\mathcal{G}_{ij} \prec \Psi$.

The above arguments show that the counterparts of Lemmas 5.9 and 5.10 of [71] hold. Therefore, by the same arguments as in Lemma 5.12 of [71], we can conclude the proof.

□

Next we provide some useful controls whose proofs and results will be used in the proof of Lemmas S.2.5 and S.2.6. The results and arguments are analogous to Lemma 5.4 of [56]. We only point out the key ingredients in our proof and refer the readers to [56] for more details.

LEMMA S.2.8. *Suppose the assumptions of Theorem S.1.9 hold. Then we have that on Ω and for all $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$*

$$\operatorname{Im} m_1(z) \prec \frac{1}{n\eta_0}, \quad \operatorname{Im} m_Q(z) \prec \frac{1}{n\eta_0}.$$

PROOF. Due to similarity, we focus our proof on $\operatorname{Im} m_Q(z)$. We prove by contradiction. Given some $\epsilon > 0$, conditional on Ω, for some sufficiently small constants $0 < c_1, c_2 < 1$, we first introduce a probability event $\Xi_1 \equiv \Xi_1(\epsilon)$ so that the followings holds:

1. For $z \in \mathbf{D}_b'$,

$$|m_Q(z) - m_n(z)| + |m_1(z) - m_{1n}(z)| + |m_2(z) - m_{2n}(z)| \leqslant (n\eta)^{-1/4 + c_1\epsilon}.$$

2. For $z \in \mathbf{D}_b'$ and Z_i in (S.9),

$$\max_i Z_i \leqslant n^{c_2\epsilon}\Psi.$$

According to Proposition S.2.7 and (S.49), we find that there exists some large constant $D > 0$ so that $\mathbb{P}(\Xi_1) = 1 - n^{-D}$. In what follows, we restrict ourselves on Ξ_1 so that the discussions are purely deterministic.

Assuming that

$$\operatorname{Im} m_1(z) > n^{\epsilon}\frac{1}{n\eta_0}.$$

We then conclude from the definition of Ψ in (S.48) that

$$\Psi = \mathrm{o}(\operatorname{Im} m_1(z)). \tag{S.54}$$

Moreover, by (S.19) and (S.20), we readily see that $\operatorname{Im} m_{1n}(z) \ll \operatorname{Im} m_1(z)$. This implies that for some constant $C > 0$

$$|m_{1n}(z) - m_1(z)| \geq |\operatorname{Im} m_{1n} - \operatorname{Im} m_1| > Cn^{\epsilon}\frac{1}{n\eta_0}. \tag{S.55}$$

On the other hand, by Proposition S.2.7 and Assumption S.1.1, we see that (S.53) still holds. Together with m_{1n} in (6.2), using (S.9), we find that

$$\begin{aligned} m_{1n} - m_1 &= \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}-m_1)+\xi_j^2 \mathrm{O}(n^{c_2\epsilon}\Psi)}{(1+\xi_j^2 m_{1n})(1+\xi_j^2 m_1 + \mathrm{O}(n^{c_2\epsilon}\Psi))}}{(z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})(z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_1 + \mathrm{O}(n^{c_2\epsilon}\Psi)})} + \mathrm{O}(n^{c_2\epsilon}\Psi) \\ &= \mathsf{C}_1(m_{1n} - m_1) + \mathsf{C}_2 + \mathrm{O}(n^{c_2\epsilon}\Psi), \end{aligned} \tag{S.56}$$

where $\mathsf{C}_1, \mathsf{C}_2$ are defined as

$$\mathsf{C}_1 := \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{(1+\xi_j^2 m_{1n})(1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi))}}{(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi)})},$$

$$\mathsf{C}_2 := \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^2 \mathrm{O}(n^{c_2\epsilon}\Psi)}{(1+\xi_j^2 m_{1n})(1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi))}}{(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi)})}.$$

We first control C_2. It is easy to see that $m_1 \sim 1$ by contradiction. If $m_1 \ll 1$, one can observe from (S.50) that $m_2 \sim 1$ which yields $m_1 \asymp 1$ by (S.53). If $m_1 \gg 1$, we have $m_2 \ll 1$ from (S.50), and then it gives that $m_1 \asymp 1$ by (S.53). Similarly, we can show that $m_2 \asymp 1$. Together with Proposition S.2.7, we find that $m_{1n}, m_{2n} \asymp 1$. Since $z \asymp 1$, using the definition of m_{2n} in (6.2) and Proposition S.2.7, we find that

$$\frac{1}{n}\sum_j \frac{\xi_j^2 \mathrm{O}(n^{c_2\epsilon}\Psi)}{(1+\xi_j^2 m_{1n})(1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi))} = \mathrm{O}(n^{c_2\epsilon}\Psi).$$

Moreover, by Proposition S.2.7, (S.46) and Assumption S.1.1, we find that

$$\frac{1}{n}\sum_i \frac{1}{(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})(z-\frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_1+\mathrm{O}(n^{c_2\epsilon}\Psi)})} \asymp 1.$$

This yields that

$$\mathsf{C}_2 = \mathrm{O}(n^{c_2\epsilon}\Psi). \tag{S.57}$$

For C_1, using Proposition S.2.7 and Assumption S.1.1, by an argument similar to (S.42), we can conclude that when n is sufficiently large, for some constant $0 < \mathfrak{c} < 1$,

$$\mathsf{C}_1 \leqslant \mathfrak{c}. \tag{S.58}$$

Combining (S.56), (S.57) and (S.58), we conclude that

$$|m_{1n} - m_1| = \mathrm{O}(n^{c_2\epsilon}\Psi),$$

which contradicts (S.55) since $c_2 < 1$ is sufficiently small. This completes our proof for each fixed z. For uniformity in z, we can follow a standard lattice argument as discussed below (S.47). This finishes the proof. The discussion for m_Q follows from an analogous discussion with the help of (S.50) and (S.53). □

REMARK S.2.9. Two remarks are in order. First, it is easy to see that repeating the proof of Lemma S.2.8, we can prove the results for all η as specified in (S.6). Second, we remark that combining (S.49) and Lemma S.2.8, when $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$, we have that conditional on Ω

$$Z_i \prec \frac{1}{n\eta_0}.$$

Armed with the above discussions and results, following the strategies of Lemmas 5.6 and 5.7 of [56] or [52], we prove Lemmas S.2.5 and S.2.6 using similar arguments as in Lemma S.2.8. Due to similarity, we only provide the key ingredients.

Proof of Lemmas S.2.5 and S.2.6. Due to similarity, we focus our proof on Lemma S.2.5 and briefly mention that of Lemma S.2.6 in the end. Due to similarity, we only explain $|m_{1n}(z)-m_1(z)|$.

The proof is similar to that of Lemma S.2.8 and we prove by contradiction. We also restrict ourselves on the event Ξ_1 in Lemma S.2.8. We assume that $|m_{1n}(z)-m_1(z)| > n^{\epsilon}(n\eta_0)^{-1}$. To see a contraction, in addition to the the arguments of Lemma S.2.8, we need to provide a finer control for Ψ since in the current proof it depends on η instead of η_0. Note that

$$
\begin{aligned}
n^{c_2\epsilon}\Psi &= n^{c_2\epsilon}\sqrt{\frac{|\operatorname{Im} m_1 - \operatorname{Im} m_{1n} + \operatorname{Im} m_{1n}|}{n\eta}} + n^{c_2\epsilon}\frac{1}{n\eta} \\
&\leq n^{c_2\epsilon}\sqrt{\frac{|\operatorname{Im} m_1 - \operatorname{Im} m_{1n}|}{n\eta}} + n^{c_2\epsilon}\sqrt{\frac{\operatorname{Im} m_{1n}}{n\eta}} + n^{c_2\epsilon}\frac{1}{n\eta} \\
&= \mathrm{o}(|m_1 - m_{1n}|),
\end{aligned} \tag{S.59}
$$

where in the last step we used (S.18) and the assumption $|m_{1n}(z)-m_1(z)| > n^{\epsilon}(n\eta_0)^{-1} \gg (n\eta)^{-1}$ when $n^{-1/2+\epsilon_d} \leqslant \eta \leqslant n^{-1/(d+1)+\epsilon_d}$. Replacing $n^{c_2\epsilon}\Psi$ with $\mathrm{o}(|m_1-m_{1n}|)$ in the arguments between (S.56) and (S.58), we find that $|m_{1n}-m_1| = \mathrm{o}(|m_{1n}-m_1|)$ which is a contraction. This proves the result for each fixed z. The uniformity follows from the same lattice argument as mentioned in the end of the proof of Lemma S.2.8.

The proof of Lemma S.2.6 is similar. We also prove by contradiction and assume that $n^{\epsilon}(n\eta_0)^{-1} < |m_1(z)-m_{1n}(z)| \leq n^{\epsilon+\epsilon_d}(n\eta_0)^{-1}$. Under this assumption, together with (S.19) and (S.20), we find that (S.59) still holds true. The rest of the arguments are similar and we omit the details.

□

APPENDIX S.3: LOCATIONS FOR EXTREME BOOTSTRAPPED EIGENVALUES AND PROOF OF THE MAIN RESULTS

In this section, we study the first order convergent limits of the largest eigenvalues of Q, i.e., $\lambda_1(Q)$. In Section S.3.1, we investigate the case when $\{\xi_i^2\}$ have unbounded support as in (i) of Assumption 2.2. In Section S.3.2, we study the bounded support case as in (ii) of Assumption 2.2.

S.3.1. The unbounded support case

In order to quantify the location of $\lambda_1 \equiv \lambda_1(Q)$, we need to introduce several auxiliary quantities. Recall ϑ_1 defined in (6.5). Similarly, we denote ϑ_2 by replacing $\xi_{(1)}^2$ with $\xi_{(2)}^2$ in (6.5). Moreover, for d_1 and the sufficiently small constant $\epsilon > 0$ in (S.4), we denote

$$
\vartheta_1^{\pm} := \vartheta_1 \pm n^{-1/2+2\epsilon} d_1, \tag{S.1}
$$

and recall that

$$
Q^{(1)} := Q - \mathbf{y}_{(1)}\mathbf{y}_{(1)}^*, \tag{S.2}
$$

where $\mathbf{y}_{(1)}$ is the column of Y in (2.1) associated with $\xi_{(1)}^2$. Accordingly, we denote the largest eigenvalue of $Q^{(1)}$ as $\lambda_1^{(1)} \equiv \lambda_1(Q^{(1)})$. Throughout this section, we shall prove Figure S.3 so that the location of λ_1 can be quantified with high probability on the event Ω.

More formally, the main result is summarized in Proposition S.3.1 below.

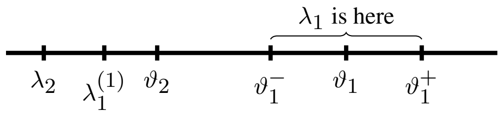


FIG S.3. *Location of the largest eigenvalue of Q.*

PROPOSITION S.3.1. *Suppose Assumptions 2.1, 2.4 and (i) of Assumption 2.2 hold. For some sufficiently small constant $\epsilon > 0$ and $\vartheta_1^{\pm}$ defined in (S.1), condition on the probability event Ω in Lemma S.1.12, with high probability, we have that*

$$\lambda_1 \in [\vartheta_1^-, \vartheta_1^+].$$

We now proceed to the proof of Proposition S.3.1 following the structure described in Figure S.3.

Proof of Proposition S.3.1. In what follows, we restrict the discussion on the probability event Ω in Lemma S.1.12. By Weyl's inequality, we have that $\lambda_2 \leqslant \lambda_1^{(1)}$. Moreover, by (S.36), we see that with high probability $\lambda_1^{(1)} < \vartheta_2$. The rest of the proof leaves to prove that the following two claims:

$$\vartheta_1 - \vartheta_2 \geqslant n^{1/\alpha} \log^{-1} n, \tag{S.3}$$

and for $Q^{(1)}$ in (S.2) and

$$M(\lambda) = 1 + \mathbf{y}_{(1)}^* G_1^{(1)}(\lambda) \mathbf{y}_{(1)}, \ G_1^{(1)}(\lambda) := (Q^{(1)} - \lambda I)^{-1}, \tag{S.4}$$

$M(\lambda)$ changes sign with high probability at ϑ_1^- and ϑ_1^+. In fact, for λ_1, it should satisfy the following equation with high probability

$$\det(\lambda_1 I - \mathbf{y}_{(1)} \mathbf{y}_{(1)}^* - Q^{(1)}) = 0 \Rightarrow M(\lambda_1) = 0, \tag{S.5}$$

as long as $\lambda_1 > \lambda_1^{(1)}$. On the other hand, if $M(\lambda)$ changes sign at $\vartheta_1^{\pm}$, by continuity, there must at least be an eigenvalue of Q in the interval $[\vartheta_1^-, \vartheta_1^+]$. If (S.3) holds, combining the above arguments, we see that the only possibility is λ_1 and it is also true that $\lambda_1 > \lambda_1^{(1)}$.

We first justify (S.3). Recall that ϑ_1 is defined in (6.5) according to

$$1 + (\xi_{(1)}^2 + d_1) m_{1n}(\vartheta_1) = 0.$$

Together with (S.2) and (S.3), we readily obtain that

$$1 = \frac{1}{n} \sum_{i=1}^p \frac{\sigma_i}{\frac{\vartheta_1}{\xi_{(1)}^2 + d_1} - \frac{\sigma_i}{n} \sum_{j=1}^n \frac{\xi_{(j)}^2}{\xi_{(1)}^2 + d_1 - \xi_{(j)}^2}}. \tag{S.6}$$

Recall (S.9). Using the definition of d_1 and (S.25), we see that on Ω, for some constant $C > 0$

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \frac{\xi_{(j)}^2}{\xi_{(1)}^2 + d_1 - \xi_{(j)}^2} &= \frac{1}{n} \frac{\xi_{(1)}^2}{d_1} + \frac{1}{n} \sum_{j=2}^n \frac{\xi_{(j)}^2}{\xi_{(1)}^2 + d_1 - \xi_{(j)}^2} \\ &\leq \frac{C n^{\epsilon} \log n}{n} + \frac{1}{n} \sum_{j=2}^n \frac{\xi_{(j)}^2}{\xi_{(1)}^2 + d_1 - \xi_{(2)}^2} \\ &\leq \frac{C n^{\epsilon} \log n}{n} + \frac{C \log n}{n^{1/\alpha}} = C e. \end{aligned} \tag{S.7}$$

Recall the definition of φ in Theorem 3.1, the above arguments imply that on Ω

$$\frac{\vartheta_1}{\xi_{(1)}^2+d_1}=\varphi+\mathrm{O}(e). \tag{S.8}$$

Moreover, after some calculation from (S.6), one has

$$\frac{\vartheta_1}{\xi_{(1)}^2+d_1}-\varphi=\frac{\xi_{(1)}^2+d_1}{\vartheta_1}\times\frac{1}{n}\sum_{i=1}^n\sigma_i^2\times\frac{1}{n}\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(1)}^2-\xi_{(j)}^2}+\mathrm{O}(e^2).$$

Then by Assumption 2.2, we can obtain that

$$\frac{1}{n}\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(2)}^2-\xi_{(j)}^2}=\frac{\mathbb{E}\xi^2}{\xi_{(1)}^2}+\mathrm{O}(e^2).$$

Therefore, we conclude a enhanced form for (S.8) that

$$\vartheta_1=\varphi\times(\xi_{(1)}^2+d_1)+\mathbb{E}\xi^2\times\frac{1}{n\varphi}\sum_{i=1}^n\sigma_i^2+\mathrm{O}(e). \tag{S.9}$$

By an analogous argument, we have that for some constants $C_k>0, k=1,2,3$,

$$\begin{aligned}
&\frac{1}{n}\Big(\sum_{j=1}^n\frac{\xi_{(j)}^2}{\xi_{(1)}^2+d_1-\xi_{(j)}^2}-\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(2)}^2+d_1-\xi_{(j)}^2}\Big)\\
&=\frac{1}{n}\Big(\frac{\xi_{(1)}^2}{d_1}+\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(1)}^2+d_1-\xi_{(j)}^2}-\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(2)}^2+d_1-\xi_{(j)}^2}\Big)\\
&=C_1e-\frac{1}{n}\sum_{j=2}^n\frac{\xi_{(j)}^2(\xi_{(1)}^2-\xi_{(2)}^2)}{(\xi_{(1)}^2+d_1-\xi_{(j)}^2)(\xi_{(2)}^2+d_1-\xi_{(j)}^2)}\\
&\geq C_1e-\frac{1}{n}\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(2)}^2+d_1-\xi_{(j)}^2}\\
&\geqslant C_2e,
\end{aligned}$$

and on the other hand

$$\begin{aligned}
&\frac{1}{n}\Big(\sum_{j=1}^n\frac{\xi_{(j)}^2}{\xi_{(1)}^2+d_1-\xi_{(j)}^2}-\sum_{j=2}^n\frac{\xi_{(j)}^2}{\xi_{(2)}^2+d_1-\xi_{(j)}^2}\Big)\\
&\leq\frac{1}{n}\Big(\frac{\xi_{(1)}^2}{d_1}+\sum_{j=2}\frac{\xi_{(j)}^2}{\xi_{(1)}^2+d_1-\xi_{(j)}^2}+\sum_{j=2}\frac{\xi_{(j)}^2}{\xi_{(2)}^2+d_1-\xi_{(j)}^2}\Big)\\
&\leqslant C_3e.
\end{aligned}$$

Using the above control, the definition of ϑ_2 and a discussion similar to (S.8), we can prove that

$$\frac{\vartheta_2}{\xi_{(2)}^2+d_1}=\phi\bar{\sigma}+\mathrm{O}(e). \tag{S.10}$$

Combining (S.8) and (S.10), we immediately see that

$$\frac{\vartheta_1}{\xi_{(1)}^2+d_1}-\frac{\vartheta_2}{\xi_{(2)}^2+d_1}=\mathrm{O}(e). \tag{S.11}$$

This implies that

$$\begin{aligned}\vartheta_1-\vartheta_2&=(\xi_{(1)}^2+d_1)\Big(\frac{\vartheta_1}{\xi_{(1)}^2+d_1}-\frac{\vartheta_2}{\xi_{(2)}^2+d_1}\Big)+\vartheta_2\Big(\frac{\xi_{(1)}^2+d_1}{\xi_{(2)}^2+d_1}-1\Big)\\&=(\xi_{(1)}^2+d_1)\Big(\frac{\vartheta_1}{\xi_{(1)}^2+d_1}-\frac{\vartheta_2}{\xi_{(2)}^2+d_1}\Big)+\frac{\vartheta_2}{\xi_{(2)}^2+d_1}(\xi_{(1)}^2-\xi_{(2)}^2)\\&\geq n^{2/\alpha}\log^{-1}n,\end{aligned}$$

where in the third step we used (S.11), (S.25) and the definition of e_2 in (S.9). This completes the proof of (S.3).

Next, we will show that

$$M(\vartheta_1^-)<0,\ M(\vartheta_1^+)>0.$$

Due to similarity, in what follows, we focus on the first inequality. Note that

$$\mathbf{y}_1^*G_1^{(1)}(\vartheta_1^-)\mathbf{y}_1=\xi_{(1)}^2\mathbf{u}_1^*\Sigma^{1/2}G_1^{(1)}(\vartheta_1^-)\Sigma^{1/2}\mathbf{u}_1.$$

Moreover, recall that $m_1^{(1)}(\vartheta_1^-)=n^{-1}\operatorname{tr}(\Sigma^{1/2}G_1^{(1)}(\vartheta_1^-)\Sigma^{1/2})$. Then according to Lemma S.1.15, we have that

$$\begin{aligned}\mathbf{y}_1^*G_1^{(1)}(\vartheta_1^-)\mathbf{y}_1&=\xi_{(1)}^2m_1^{(1)}(\vartheta_1^-)+\mathrm{O}_\prec\Big(\frac{\xi_{(1)}^2}{n}\|G_1^{(1)}(\vartheta_1^-)\|_F\Big)\\&=\xi_{(1)}^2m_1^{(1)}(\vartheta_1^-)+\mathrm{O}_\prec\Big(\frac{\xi_{(1)}^2}{n^{1/2+1/\alpha}}\Big),\end{aligned}\tag{S.12}$$

where in the second step we used (S.3) and the fact $\vartheta_2>\lambda_1^{(1)}$. Moreover, for some sufficiently small constant $\epsilon_0>0$ and $z_0=\vartheta_1^-+\mathrm{i}n^{-1/2-\epsilon_0}$, we can decompose that

$$\begin{aligned}m_1^{(1)}(\vartheta_1^-)&=\Big[m_1^{(1)}(\vartheta_1^-)-m_1^{(1)}(z_0)\Big]+\Big[m_1^{(1)}(z_0)-m_{1n}^{(1)}(z_0)\Big]+\Big[m_{1n}^{(1)}(z_0)-m_{1n}^{(1)}(\vartheta_1^-)\Big]+m_{1n}^{(1)}(\vartheta_1^-)\\&=\mathsf{P}_1+\mathsf{P}_2+\mathsf{P}_3+m_{1n}^{(1)}(\vartheta_1^-).\end{aligned}\tag{S.13}$$

First, by Theorem S.1.8, we have that $\mathsf{P}_2\prec n^{-1/2-2/\alpha}$. Second, let $\{\mathbf{v}_i^{(1)}\}$ be the eigenvectors of $Q^{(1)}$ associated with the eigenvalues $\{\lambda_i^{(1)}\}$, then we have that

$$\begin{aligned}|\mathsf{P}_1|&\leq\frac{1}{n}\sum_{i=1}^p|T\mathbf{v}_i^{(1)}|^2\left|\frac{1}{\lambda_i^{(1)}-\vartheta_1^-}-\frac{1}{\lambda_i^{(1)}-z_0}\right|\\&=\frac{1}{n}\sum_{i=1}^p|T\mathbf{v}_i^{(1)}|^2\left|\frac{\mathrm{i}n^{-1/2-\epsilon_0}}{(\lambda_i^{(1)}-\vartheta_1^-)(\lambda_i^{(1)}-z_0)}\right|\\&\leq\frac{1}{n}\sum_{i=1}^p|T\mathbf{v}_i^{(1)}|^2\Big|\frac{\mathrm{i}n^{-1/2-\epsilon_0}+\mathrm{O}_\prec(n^{-2/\alpha-1/2}\log^2n)}{|\lambda_i^{(1)}-z_0|^2}\Big|\\&\prec\operatorname{Im}(m_1^{(1)}(z_0))\times\mathrm{O}_\prec(n^{-1/2-\epsilon_0})\prec n^{-1-1/\alpha},\end{aligned}$$

where in the third step we used (S.3) and the fact $\vartheta_2 > \lambda_1^{(1)}$ and in the last step we used Lemma S.1.2, (S.8) and (S.25). Third, according to Lemma 6.2, we can decompose that

$$
\begin{aligned}
\mathsf{P}_3 &= \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-z_0(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-z_0(1+m_{1n}^{(1)}(z_0)\xi_{(j)}^2)})} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-\vartheta_1^-(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-\vartheta_1^-(1+m_{1n}^{(1)}(\vartheta_1^-)\xi_{(j)}^2)})} \\
&= [\frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-z_0(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-z_0(1+m_{1n}^{(1)}(z_0)\xi_{(j)}^2)})} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-z_0(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-\vartheta_1^-(1+m_{1n}^{(1)}(\vartheta_1^-)\xi_{(j)}^2)})}] \\
&+ [\frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-z_0(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-\vartheta_1^-(1+m_{1n}^{(1)}(\vartheta_1^-)\xi_{(j)}^2)})} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-\vartheta_1^-(1+\frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2}{-\vartheta_1^-(1+m_{1n}^{(1)}(\vartheta_1^-)\xi_{(j)}^2)})}] \\
&:= \mathcal{M}_{31}^{(1)} + \mathcal{M}_{32}^{(1)}.
\end{aligned}
$$

Note that according to (S.8), (S.3), (S.25) and Lemma S.1.2, we conclude that with high probability $|1+\sigma_i m_{2n}^{(1)}(z_0)|, |1+\sigma_i m_{2n}^{(1)}(\vartheta_1^-)| \sim 1$. For $\mathcal{M}_{31}^{(1)}$, using the definitions in (6.2) and the above bounds, we have that with high probability

$$
\begin{aligned}
\mathcal{M}_{31}^{(1)} &= \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i^2}{-z_0(1+\sigma_i m_{2n}^{(1)}(z))(1+\sigma_i m_{2n}^{(1)}(\vartheta_1^-))}\left(m_{2n}^{(1)}(\vartheta_1^-) - m_{2n}^{(1)}(z)\right) \\
&= \mathrm{O}(|z_0|^{-1}) \times \frac{1}{n}\sum_{j=2}^n \Big(\frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))} - \frac{\xi_{(j)}^2}{-z(1+\xi_{(j)}^2 m_{1n}^{(1)}(z_0))}\Big) \\
&= \mathrm{O}(|z_0|^{-1}) \times \frac{1}{n}\sum_{j=2}^n \Big(\frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))} - \frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(z_0))}\Big) \\
&+ \mathrm{O}(|z_0|^{-1}) \times \frac{1}{n}\sum_{j=2}^n \Big(\frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(z))} - \frac{\xi_{(j)}^2}{-z(1+\xi_{(j)}^2 m_{1n}^{(1)}(z_0))}\Big) \\
&= \mathrm{O}(|z_0|^{-1}) \times \frac{1}{n}\sum_{j=2}^n \frac{\xi_{(j)}^4 (m_{1n}^{(1)}(z_0) - m_{1n}^{(1)}(\vartheta_1^-))}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))(1+\xi_{(j)}^2 m_{1n}^{(1)}(z_0))} + \mathrm{O}(|z_0|^{-1}) \times \frac{1}{n}\sum_{j=2}^n \frac{\xi_{(j)}^2(\vartheta_1^- - z_0)}{z\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(z_0))} \\
&= \mathrm{o}(1) \times (m_{1n}^{(1)}(z_0) - m_{1n}^{(1)}(\vartheta_1^-)) + \mathrm{O}(n^{-2/\alpha - 1/2 - \epsilon_0}).
\end{aligned} \tag{S.14}
$$

where in the second to last step we used Lemma S.1.2 and in the last step we used Lemma S.1.2 and (S.25). This implies with high probability

$$
\mathcal{M}_{31}^{(1)} = \mathrm{O}\left(n^{-2/\alpha - 1/2 - \epsilon_0}\right).
$$

Similarly, for $\mathcal{M}_{32}^{(1)}$, we have that

$$
\begin{aligned}
\mathcal{M}_{32}^{(1)} &= \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i(z_0 - \vartheta_1^-)}{z_0\vartheta_1^-(1+\sigma_i m_{2n}^{(1)}(\vartheta_1^-))} \\
&= \mathrm{O}(n^{-2/\alpha - 1/2 - \epsilon_0}).
\end{aligned} \tag{S.15}
$$

Combining the above arguments, we have that $\mathsf{P}_3 = \mathrm{O}\left(n^{-2/\alpha - 1/2 - \epsilon_0}\right)$.

Inserting the bounds of $\mathsf{P}_k, 1 \leqslant k \leqslant 3$ into (S.13), we conclude that

$$|m_1^{(1)}(\vartheta_1^-) - m_{1n}^{(1)}(\vartheta_1^-)| \prec n^{-1/\alpha - 1/2 - \epsilon_0}.$$

Together with (S.12) and (S.25), it yields that

$$M(\vartheta_1^-) = 1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1^-) + \mathrm{O}_{\prec}(n^{-1/2-\epsilon_0}). \tag{S.16}$$

In what follows, we study $1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1^-)$. We rewrite that

$$1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1^-) = 1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1) - (\xi_{(1)}^2 + d_1)\big(m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\vartheta_1^-)\big). \tag{S.17}$$

Using the definition for ϑ_1 that $1 + (\xi_{(1)}^2 + d_1) m_{1n}(\vartheta_1) = 0$ and the definitions in (6.2), by a discussion similar to (S.14), we have that for some constant $C > 0$

$$\begin{aligned}
&1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1) \\
&= 1 + (\xi_{(1)}^2 + d_1) m_{1n}(\vartheta_1) + (\xi_{(1)}^2 + d_1)\big(m_{1n}^{(1)}(\vartheta_1) - m_{1n}(\vartheta_1)\big) \\
&= (\xi_{(1)}^2 + d_1) \frac{1}{n} \sum_{i=1}^{p} \Big(\frac{\sigma_i}{-\vartheta_1(1 + \sigma_i m_{2n}^{(1)}(\vartheta_1))} - \frac{\sigma_i}{-\vartheta_1(1 + \sigma_i m_{2n}(\vartheta_1))} \Big) \\
&= (\xi_{(1)}^2 + d_1) \left[\frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i^2}{-\vartheta_1(1 + \sigma_i m_{2n}^{(1)}(\vartheta_1))(1 + \sigma_i m_{2n}(\vartheta_1))} \right] \big(m_{2n}(\vartheta_1) - m_{2n}^{(1)}(\vartheta_1)\big) \\
&\leqslant C \frac{(\xi_{(1)}^2 + d_1)}{\vartheta_1} \times \Big| \frac{1}{n} \sum_{j=1}^{n} \frac{\xi_{(j)}^2}{-\vartheta_1(1 + \xi_{(j)}^2 m_{1n}(\vartheta_1))} - \frac{1}{n} \sum_{j=2}^{n} \frac{\xi_{(j)}^2}{-\vartheta_1(1 + \xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1))} \Big| \\
&\leqslant C \frac{(\xi_{(1)}^2 + d_1)}{\vartheta_1} \times \xi_{(2)}^2 |m_{1n}^{(1)}(\vartheta_1) - m_{1n}(\vartheta_1)| + n^{-1}.
\end{aligned} \tag{S.18}$$

where in the last step we again used (S.25). This yields that

$$|(\xi_{(1)}^2 + d_1)\big(m_{1n}^{(1)}(\vartheta_1) - m_{1n}(\vartheta_1)\big)| \leqslant C \frac{(\xi_{(1)}^2 + d_1)}{\vartheta_1} \times \xi_{(2)}^2 |m_{1n}^{(1)}(\vartheta_1) - m_{1n}(\vartheta_1)| + n^{-1},$$

which implies that $(\xi_{(1)}^2 + d_1)\big(m_{1n}^{(1)}(\vartheta_1) - m_{1n}(\vartheta_1)\big) = \mathrm{O}(n^{-1})$. Together with (S.17), we have

$$1 + (\xi_{(1)}^2 + d_1) m_{1n}^{(1)}(\vartheta_1^-) = -(\xi_{(1)}^2 + d_1)\big(m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\vartheta_1^-)\big) + \mathrm{O}(n^{-1}). \tag{S.19}$$

Recall that we have proved the facts that $\vartheta_1, \vartheta_1^- > \lambda_1^{(1)}$, by Theorem S.1.9 and the monotonicity of $m_1^{(1)}$ outside the bulk, the first term on the right-hand side of (S.19) is negative. In order to show $M(\vartheta_1^-) < 0$, in light of (S.16), it suffices to show that its magnitude is much larger than $\mathrm{O}(n^{-1/2-\epsilon_0})$. To see this, we decompose that

$$
\begin{aligned}
& m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\vartheta_1^-) \\
&= \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1(1+\sigma_i m_{2n}^{(1)}(\vartheta_1))} - \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1^-(1+\sigma_i m_{2n}^{(1)}(\vartheta_1^-))} \\
&= \Big[\frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1))})} - \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})}\Big] \\
&+ \Big[\frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})} - \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\vartheta_1^-(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})}\Big] \\
&:= \tilde{\mathcal{M}}_{11}^{(1)} + \tilde{\mathcal{M}}_{12}^{(1)}.
\end{aligned}
$$

Similar to the discussion of (S.14), we have that $\tilde{\mathcal{M}}_{11}^{(1)}$,

$$
\begin{aligned}
\tilde{\mathcal{M}}_{11}^{(1)} &= \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i^2\Big(\frac{1}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\lambda))} - \frac{1}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1))}\Big)}{-\vartheta_1(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1))})(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})} \\
&= \mathrm{O}(\frac{1}{\vartheta_1}) \times \frac{1}{n}\sum_{j=2}^{n} \frac{\xi_{(j)}^4(m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\vartheta_1^-))}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1))} \\
&= \mathrm{o}(1) \times (m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\vartheta_1^-)).
\end{aligned}
$$

Moreover, similar to (S.15), for $\tilde{\mathcal{M}}_{12}^{(1)}$ we have that with high probability

$$
\begin{aligned}
\tilde{\mathcal{M}}_{12}^{(1)} &= \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i(\vartheta_1 - \vartheta_1^-)}{\vartheta_1\vartheta_1^-(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})(1+\frac{\sigma_i}{n}\sum_{j=2}^{n}\frac{\xi_{(j)}^2}{-\vartheta_1^-(1+\xi_{(j)}^2 m_{1n}^{(1)}(\vartheta_1^-))})} \\
&\asymp \frac{\vartheta_1 - \vartheta_1^-}{\vartheta_1^-\vartheta_1} \asymp n^{-1/2-1/\alpha+\epsilon}.
\end{aligned}
$$

This implies that

$$
m_{1n}^{(1)}(\vartheta_1) - m_{1n}^{(1)}(\lambda) \asymp n^{-1/2-1/\alpha+\epsilon}.
$$

Together with (S.19), the definition of d_2 and (S.25), we readily see that

$$
1 + (\xi_{(1)}^2 + d_2) m_{1n}^{(1)}(\vartheta_1^-) \asymp -n^{-1/2+\epsilon}, \tag{S.20}
$$

which concludes the proof of $M(\vartheta_1^-) < 0$ when n is sufficiently large. Similarly, we can prove that $M(\vartheta_1^+) > 0$. This completes the proof.

□

S.3.2. The bounded support case

In this section, we study the first order convergence of λ_1 under the assumptions of part (1) of Theorem 3.3. The other parts will be discussed in Section S.4. The main result of this section can be summarized in the following proposition.

PROPOSITION S.3.2. *Suppose the assumptions of part (1) of Theorem 3.3 hold, then conditional on the probability event as in Lemma S.1.5, we have that*

$$\left|\lambda_1-\left(\widehat{L}_+-\frac{1-\phi\widehat{\mathsf{s}}_3}{\widehat{\mathsf{s}}_4}\frac{l-\xi_{(1)}^2}{l\xi_{(1)}^2}\right)\right|=\mathrm{O}_{\mathbb{P}}\left[\frac{1}{n^{1/(d+1)}}\left(\frac{n^{3\epsilon_d}}{n^{-1/(d+1)+1/2}}+\frac{\log n}{n^{1/(d+1)}}\right)\right].$$

The proof of Proposition S.3.2 relies crucially on the following lemma whose justification will be provided in the end of this section.

LEMMA S.3.3. *Suppose the assumptions of Proposition S.3.2 hold. Recall E_0 defined in (6.7). Conditional on the probability event Ω as in Lemma S.1.5, we have that*

$$\lambda_1=E_0+\mathrm{O}_{\mathbb{P}}\left(n^{-1/2+3\epsilon_d}\right), \tag{S.21}$$

and

$$\mathrm{Re}\, m_{1n}(\lambda_1+\mathrm{i}\eta_0)=-\frac{1}{\xi_{(1)}^2}+\mathrm{O}_{\mathbb{P}}(n^{-1/2+3\epsilon_d}). \tag{S.22}$$

Armed with Lemma S.3.3, we proceed to the proof of Proposition S.3.2.

Proof of Proposition S.3.2. Recall (S.13), we have that conditional on Ω, $m_{1n}(\widehat{L}_+)=-l^{-1}$. Together with (S.16), we conclude that

$$\mathrm{Re}\, m_{1n}(\widehat{L}_++\mathrm{i}n^{-1/2-\epsilon_d})=-l^{-1}+\mathrm{O}(n^{-1/2-\epsilon_d}).$$

Moreover, according to the definition in (6.7), we have that

$$\mathrm{Re}\, m_{1n}(E_0+\mathrm{i}n^{-1/2-\epsilon_d})=-\xi_{(1)}^2.$$

By (S.27), we obtain that conditional on Ω

$$\mathrm{Re}\, m_{1n}(\widehat{L}_++\mathrm{i}n^{-1/2-\epsilon_d})=\mathrm{Re}\, m_{1n}(E_0+\mathrm{i}n^{-1/2-\epsilon_d})+\mathrm{O}\left(\frac{\log n}{n^{d+1}}\right),$$

which implies that $\widehat{L}_+=E_0+\mathrm{O}(\log n/n^{d+1})$. Let Ξ be the probability event that (S.21) holds. We therefore conclude from (S.15) that when restricted on Ξ and n is sufficiently large, $\lambda_1+\mathrm{i}\eta_0\in\mathbf{D}_b$. Consequently, by part I of Lemma S.1.5, we find the following holds on Ξ

$$m_{1n}(\widehat{L}_+)-m_{1n}(\lambda_1+\mathrm{i}\eta_0)=\frac{\widehat{\mathsf{s}}_4}{(1-\phi\widehat{\mathsf{s}}_3)}\left(\widehat{L}_+-\lambda_1-\mathrm{i}\eta_0\right)+\mathrm{O}\left((\log n)(n^{-1/(d+1)})^{\min\{d,2\}}\right). \tag{S.23}$$

Again by $m_{1n}(\widehat{L}_+)=-l^{-1}$, Together with the second part of Lemma S.3.3, considering the real parts of both sides of (S.23), we obtain that on Ξ

$$-l^{-1}+\xi_{(1)}^{-2}+\mathrm{O}_{\mathbb{P}}(n^{-1/2+3\epsilon_d})=\frac{\widehat{\mathsf{s}}_4}{(1-\phi\widehat{\mathsf{s}}_3)}\left(\widehat{L}_+-\lambda_1\right)+\mathrm{O}\left((\log n)(n^{-1/(d+1)})^{\min\{d,2\}}\right).$$

This completes our proof.

□

The rest of this section leaves to the proof of Lemma S.3.3. We first prove the following lemma which will be used in the proof of Lemma S.3.3. It essentially locates the points in $\mathbf{D}_b'$ for which $\mathrm{Im}\, m_Q(z)\gg\eta_0$ near the edge. It is a counterpart of [56, Lemmas 5.12, 5.13 and 5.15] and [52, Lemmas 5.13, 5.14 and 5.16]. Due to similarity, we only sketch the key points of the proof. Recall z_0 defined in (6.7).

LEMMA S.3.4. *Suppose the assumptions of Lemma S.3.3, we have that the followings holds with high probability*

(1). *For any* $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$ *satisfying that* $|z - z_0| \geqslant n^{-1/2+3\epsilon_d}$,

$$\operatorname{Im} m_1(z) \asymp \eta_0, \ \operatorname{Im} m_Q(z) \asymp \eta_0. \tag{S.24}$$

(2). *For* $m_1^{(1)}(z)$ *and* $m_Q^{(1)}(z)$ *defined around (S.8), we have that for all* $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$,

$$\operatorname{Im} m_1^{(1)}(z) \asymp \eta_0, \ \operatorname{Im} m_Q^{(1)}(z) \asymp \eta_0. \tag{S.25}$$

(3). *There exists some* $E_0' \in \mathbb{R}$ *such that for* $z_0' = E_0' + \mathrm{i}\eta_0$, *the followings holds simultaneously*

$$|z_0' - z_0| \leqslant n^{-1/2+3\epsilon_d}, \ \textit{and} \ \operatorname{Im} m(z_0') \gg \eta_0. \tag{S.26}$$

PROOF. Due to similarity, we focus our discussion on $m_1(z)$ and will explain the minor differences for $m_Q(z)$ from line to line. Recall (S.9). Our proof relies on the following fluctuation average which provides a stronger control on $n^{-1}\sum_{i=1}^p Z_i$ than the one in Remark S.2.9. They are counterparts of Lemmas 5.8 and 5.9 and Corollary 5.10 of [56]. We deter its proof to Appendix S.6.3.

LEMMA S.3.5. *Suppose the assumptions of Lemma S.3.4 hold, we have that the followings holds on* Ω

(1). *For all* $i \neq 1$ *and all* $z = E + \mathrm{i}\eta_0 \in \mathbf{D}_b'$, *we have that*

$$|m_2 - m_2^{(1)}| \prec \frac{1}{n\eta_0}, \ \ |m_2 - m_2^{(i)}| + |m_2^{(i)} - m_2^{(i1)}| \prec n^{-1+1/(d+1)+4\epsilon_d}. \tag{S.27}$$

(2). *For all* $z \in \mathbf{D}_b'$,

$$\left|\frac{1}{n}\sum_{i=2}^n Z_i\right| + \left|\frac{1}{n}\sum_{i=2}^n Z_i^{(1)}\right| \prec n^{-1/2-\frac{1}{2}(\frac{1}{2}-\frac{1}{d+1})+2\epsilon_d}.$$

(3). *For all* $z \in \mathbf{D}_b'$,

$$\left|\frac{1}{n}\sum_{i=2}^n \frac{(\xi_i^2+\xi_i^4)Z_i}{(1+\xi_i^2 m_{1n}(z))^2}\right| \prec n^{-1/2-\frac{1}{2}(\frac{1}{2}-\frac{1}{d+1})+2\epsilon_d}.$$

Armed with Lemma S.3.5, we proceed to finish our proof. The proof of part (1) is similar to that of (S.20) by using the local law Theorem S.1.9. We only provide the key arguments. Using (S.51), (S.49), and (S.50), we see that

$$m_1 = -\frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_1 + Z_j}} + \frac{1}{n}\operatorname{tr}(R_1\Sigma) + \frac{1}{n}\operatorname{tr}(R_2\Sigma).$$

According to Theorem S.1.9, by a discussion similar to (S.53) using Remark S.2.9, we have that

$$\begin{aligned}
\operatorname{Im} m_1(z) &= \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i\eta_0}{|z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}+Z_j}|^2} + \frac{1}{n}\sum_{i=1}^p \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n \frac{\xi_j^4 \operatorname{Im} m_1 + \xi_j^2 \operatorname{Im} Z_j}{|1+\xi_j^2 m_{1n}+Z_j|^2}}{|z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}+Z_j}|^2} \\
&\quad + \mathrm{O}_\prec\left(\frac{1}{n}\sum_{j=1}^n \frac{Z_j}{|1+\xi_j^2 m_{1n}+Z_j|^2} + \frac{1}{(n\eta_0)^2}\right) \\
&:= \mathsf{R}_1 + \mathsf{R}_2 + \mathsf{R}_3.
\end{aligned} \tag{S.28}$$

Together with Assumption S.1.1, (S.27) and Remark S.2.9, we find that for some small constant $c' > 0$, when n is sufficiently large,

$$\left| z - \frac{\sigma_i}{n} \sum_{j=1}^{n} \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z) + Z_j} \right| \geqslant c'. \tag{S.29}$$

This implies that

$$\mathsf{R}_1 \asymp \eta_0. \tag{S.30}$$

For R_2, on the one hand, by Theorem S.1.9 and (S.45), we can conclude that there exists some constant $0 < \mathsf{c}' < 1$,

$$\frac{1}{n} \sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n} \sum_{j=1}^{n} \frac{\xi_j^4}{|1+\xi_j^2 m_{1n} + Z_j|^2}}{|z - \frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n} + Z_j}|^2} \leqslant \mathsf{c}' < 1.$$

Moreover, as $|z - z_0| \geqslant n^{-1/2+3\epsilon_d}$, according to (S.17) and Remark S.2.9, we find that $1 + \xi_j^2 m_{1n}(z) + Z_j \geqslant C n^{-1/2+3\epsilon_d}$ for some constant $C > 0$. Together with (S.29) and (3) of Lemma S.3.5, we find that with high probability

$$\frac{1}{n} \sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n} \sum_{j=1}^{n} \frac{\xi_j^2 \operatorname{Im} Z_j}{|1+\xi_j^2 m_{1n} + Z_j|^2}}{|z - \frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n} + Z_j}|^2} \ll \eta_0.$$

Consequently, we have that with high probability

$$\mathsf{R}_2 = \mathsf{c}' \operatorname{Im} m_1 + \mathrm{o}(\eta_0). \tag{S.31}$$

Similarly, we can prove that with high probability $\mathsf{R}_3 = \mathrm{o}(\eta_0)$. Together with (S.30), (S.31) and (S.28), we can conclude the prove of $m_1(z)$. The discussion for $m_Q(z)$ is similar except we need to use (S.50) and (S.52).

The proof of part (2) is similar to that of part (1) using Lemma S.3.5, Theorem S.1.9 and Remark S.2.9. The idea is analogous to the proof of Lemma 5.13 of [56] or Lemma 5.14 of [52]. We omit further details.

Finally, for part (3), we find from Lemma S.1.15, (S.25) and (S.28) that for some small constant $\epsilon' < \epsilon_d/2$, with high probability,

$$|Z_1| \leqslant n^{-1/2+\epsilon'}. \tag{S.32}$$

Without loss of generality, we assume that $\xi_1^2 \geqslant \xi_2^2 \geqslant \cdots \geqslant \xi_n^2$. On the one hand, by the definition of $m_2(z)$ and (S.9), we have

$$m_2 = \frac{\xi_1^2 \mathcal{G}_{11}}{n} + \frac{1}{n} \sum_{i=2}^{n} \frac{\xi_i^2}{-z(1+\xi_i^2 m_1^{(i)} + Z_i)}.$$

Together with Lemma S.1.13, we see that

$$\frac{1}{\xi_1^2 \mathcal{G}_{11}} = -z(1+\xi_1^2 m_1^{(1)} + Z_1). \tag{S.33}$$

Denote

$$z_0^{\pm} = z_0 \pm n^{-1/2+3\epsilon_d}.$$

Recall (6.7) that $1+\xi_1^2\operatorname{Re} m_{1n}(E_0+\mathrm{i}\eta_0)=0$. Using Theorem S.1.9, (S.29) and (S.17), together with (S.25) and Remark S.2.9, we conclude that for some constant $C>0$

$$\frac{1}{\xi_1^2\mathcal{G}_{11}(z_0^-)}\geqslant Cn^{-1/2+3\epsilon_d}, \text{ and } \frac{1}{\xi_1^2\mathcal{G}_{11}(z_0^+)}\leqslant -Cn^{-1/2+3\epsilon_d}.$$

Consequently, by continuity, we find that there exists $z_1=E_1+\mathrm{i}\eta_0$ with $E_1\in(E_0-n^{-1/2+3\epsilon_d},E_0+n^{-1/2+3\epsilon_d})$ that $\operatorname{Re}\mathcal{G}_{11}(z_1)=0$. For the choice of z_1, together with (S.33), we find that

$$|\operatorname{Im}(z_1\xi_1^2\mathcal{G}_{11}(z_1))|=\frac{1}{|\operatorname{Im} m_1^{(1)}(z_1)+\operatorname{Im} Z_1|}\geq n^{1/2-\epsilon_d/2}, \tag{S.34}$$

where we used (S.32) with the assumption $\epsilon'<\epsilon_d/2$ and (S.25). On the other hand, following lines of the proof of [56, Lemma 5.15], by a decomposition similar to (S.21) and a discussion similar to (S.28), using Lemma S.3.5, we find that

$$\operatorname{Im} m_1(z_1)\asymp \eta_0+\frac{\operatorname{Im}(\xi_1^2 z_1\mathcal{G}_{11}(z_1))}{n}.$$

Together with (S.34), we conclude that

$$\operatorname{Im} m_1(z_1)\gg\eta_0.$$

The discussion for m_Q is similar and we omit the details. This completes our proof. □

Finally, armed with Lemma S.3.4, we proceed to the proof of Lemma S.3.3. Since the details are similar to those of Proposition 4.6 of [56] or Proposition 4.7 of [52], we only provide the key ingredients.

Proof of Lemma S.3.3. We first prove (S.21). Using the spectral decomposition of Q, for $m_Q(z)$ in (6.1), we find that

$$\operatorname{Im} m_Q(E+\mathrm{i}\eta_0)=\frac{1}{n}\sum_{i=1}^n\frac{\eta_0}{(\lambda_i-E)^2+\eta_0^2}. \tag{S.35}$$

This yields that

$$\operatorname{Im} m_Q(\lambda_1+\mathrm{i}\eta_0)\geq(n\eta_0)^{-1}\gg\eta_0,$$

where we used the definition of η_0 in (6.7). It is clear that $\lambda_1=\mathrm{O}_{\mathbb{P}}(1)$. First, if $\lambda_1\in\mathbf{D}_b'$, then the proof follows directly from (S.24). Second, if $\lambda_1\notin\mathbf{D}_b'$, on the other hand, for the upper bound, by (3) of Lemma S.3.4 and (S.19), using the definition of $\mathbf{D}_b'$ in (S.12), with an argument similar to Proposition 4.7 of [56], we have that on Ω, $\lambda_1<E_0+n^{-1/2+3\epsilon_d}$ holds with $1-\mathrm{o}(1)$ probability. On the other hand, for the lower bound, we prove by contradiction. We assume that $\lambda_1<E_0(1)-n^{-1/2+3\epsilon_d}$. Then we see from (S.35) that $\operatorname{Im} m_Q(E+\mathrm{i}\eta_0)$ is a decreasing function of E on the interval $(E_0-n^{-1/2+3\epsilon_d},E_0+n^{-1/2+3\epsilon_d})$. However, from Lemma S.3.4 and its proof (recall that $z_0^-=E_0-n^{-1/2+3\epsilon_d}$), we have seen that, $\operatorname{Im} m_Q(z_0)\gg\eta_0$, $\operatorname{Im} m_Q(z_0^-)\sim\eta_0$, which is a contradiction. It implies that $\lambda_1\geq E_0(1)-n^{-1/2+3\epsilon_d}$ and completes the proof of (S.21).

Then we prove (S.22). By (S.17), (S.19) and (S.20), we find that

$$\operatorname{Re} m_{1n}(\lambda_1+\mathrm{i}\eta_0)=\operatorname{Re} m_{1n}(z_0)+\mathrm{O}_{\mathbb{P}}(n^{-1/2+3\epsilon_d})=-\frac{1}{\xi_{(1)}^2}+\mathrm{O}_{\mathbb{P}}(n^{-1/2+3\epsilon_d}),$$

where in the last step we used (6.7). This completes our proof. □

APPENDIX S.4: PROOF OF THE RESULTS OF SECTION 3

In this section, we prove Theorems 3.1 and 3.3 using the results in Sections S.3.1 and S.3.2.

S.4.1. Proof of the results in Section 3.1

Proof of Theorem 3.1. For the first part, according to (S.8), by Lemma S.1.12, when n is sufficiently large,

$$\vartheta_1 = (\xi_{(1)}^2 + d_1)\,(\phi\bar{\sigma}_1 + \mathrm{O}_{\mathbb{P}}(e))\,.$$

Together with Proposition S.3.1, we find that

$$\lambda_1 = (\xi_{(1)}^2 + d_1)\,(\phi\bar{\sigma}_1 + \mathrm{O}_{\mathbb{P}}(e)) + \mathrm{O}_{\mathbb{P}}\left(n^{-1/2+2\epsilon} d_1\right).$$

Using (S.25) and (S.26) and the definition of d_1 in (S.4), we can complete the proof for the first part.

For the second part, it follows directly from the results in Lemma S.1.16, (S.25) and (S.26). This completes our proof. □

S.4.2. Proof of the results in Section 3.2

Proof of Theorem 3.3. Due to similarity, we focus our discussion on the separable covariance i.i.d. data as in case (2) of Assumption 2.1. For part (1), (3.6) has been proved in (II) of Lemma S.1.5. For (3.7), the proofs follow from Proposition S.3.2, II of Lemma S.1.5 with the fact $d > 1$ and (S.27). Then (3.8) follows from (3.7) and Lemma S.1.16.

Then we proceed to the proof of parts (2) and (3). Following [27, Lemma 2.5], we see that conditional on Ω in Lemma S.1.12, $(\widehat{L}_+, m_{1n}(\widehat{L}_+))$ should satisfy the following systems of equations

(S.1)
$$m_{1n}(\widehat{L}_+) = \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-\widehat{L}_+ + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)}}, \quad 1 = \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{|1+\xi_j^2 m_{1n}(\widehat{L}_+)|^2}}{\left|\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)}\right|^2}.$$

Similarly, $(L_+, m_{1n,c}(L_+))$ should satisfy the following equations

(S.2)
$$m_{1n,c}(L_+) = \frac{1}{n}\sum_i \frac{\sigma_i}{-L_+ + \sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s)}, \quad 1 = \frac{1}{n}\sum_i \frac{\sigma_i^2 \int \frac{s^2}{|1+s m_{1n,c}(L_+)|^2}\mathrm{d}F(s)}{|L_+ - \sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s)|^2}.$$

Using the definitions of s_k and $\widehat{\mathsf{s}}_k, 1 \leqslant k \leqslant 3$, by (S.27) and an argument similar to II of Lemma S.1.5, when n is sufficiently large, we see that our assumption $\phi^{-1} < \mathsf{s}_3$ implies $\phi^{-1} < \widehat{\mathsf{s}}_3$ on Ω. This yields that for some constant $\delta > 0$

$$\text{(S.3)} \qquad \frac{1}{n}\sum_i \frac{\sigma_i^2 \widehat{\mathsf{s}}_1}{(\widehat{L}_+ - \sigma_i \widehat{\mathsf{s}}_2)^2} > 1 + \delta, \quad \frac{1}{n}\sum_i \frac{\sigma_i^2 \mathsf{s}_1}{(L_+ - \sigma_i \mathsf{s}_2)^2} > 1 + \delta.$$

where the first inequality is restricted on the event Ω. From now on, for notional simplicity, we always restrict ourselves on Ω so that the discussion is purely deterministic. Recall (3.5) and (S.10). Combining the second equations in (S.1) and (S.2) with (S.3), we readily obtain that

$$\text{(S.4)} \qquad m_{1n}(\widehat{L}_+) > -l^{-1}, \;\; m_{1n,c}(L_+) > -l^{-1}.$$

Together with (S.27), we have that for all $1 \leqslant j \leqslant n$ and some constant $\delta' > 0$

$$\frac{1}{\left|1+\xi_j^2 m_{1n}(\widehat{L}_+)\right|} \geqslant \delta', \ \frac{1}{|1+sm_{1n,c}(L_+)|} \geqslant \delta' \text{ for any } 0 < s \leqslant l. \tag{S.5}$$

We now proceed to the proof. The proof consists of two steps. In the first step, we prove the results assuming that

$$\left|m_{1n,c}(L_+) - m_{1n}(\widehat{L}_+)\right| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}), \ |L_+ - \widehat{L}_+| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}). \tag{S.6}$$

In the second step, we justify (S.6). We start with step one.

Step one: Under the assumption S.6, the key component of the proof is the following lemma. Denote

$$\mathsf{C}_1 := \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{\left(L_+ - \sigma_i \int \frac{s}{1+sm_{1n,c}} \mathrm{d}F(s)\right)^2},$$

and

$$\mathcal{X} := \frac{1}{n}\sum_{j=1}^{n}\Big(\frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)} - \int \frac{s}{1+sm_{1n,c}(L_+)}\mathrm{d}F(s)\Big).$$

According to Assumption S.1.1, we have that

$$\mathsf{C}_1 \asymp 1.$$

LEMMA S.4.1. *Under the assumptions of Theorem 3.3 and (S.6), we have that*

$$\mathsf{C}_1(\widehat{L}_+ - L_+) = \mathsf{C}_1\mathcal{X} + \mathrm{O}_{\mathbb{P}}(n^{-1}).$$

Armed with Lemma S.4.1, we can easily prove parts (2) and (3). Recall the definition of v in (3.5). It is clear from (S.27) $\mathcal{X} = \mathrm{O}_{\mathbb{P}}(n^{-1/2})$, and from central limit theorem that $\mathcal{X}$ is asymptotically Gaussian with variance $n^{-1}\mathsf{v}$. We decompose that

$$\lambda_1 - L_+ = \lambda_1 - \widehat{L}_+ + \widehat{L}_+ - L_+.$$

According to [24, 28] and Lemma S.1.5, we find that on Ω, $|\lambda_1 - \widehat{L}_+| \prec n^{-2/3}$ and $n^{2/3}\gamma(\lambda_1 - \widehat{L}_+)$ follows type-1 Tracy-Widom law. This concludes the general results in part (3). Moreover, for part (2), it is easy to see that when $d > 1$, by Cauchy-Schwarz inequality, v is bounded from blow so that the Gaussian part dominates the Tracy-Widom part and we hence conclude the proof.

To complete Step one, we now prove Lemma S.4.1.

Proof of Lemma S.4.1. Using the first parts in equations (S.1) and (S.2), we see that

$$\begin{aligned}
& m_{1n,c}(L_+) - m_{1n}(\widehat{L}_+) \\
&= \frac{1}{n}\sum_i \frac{\sigma_i}{\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)}} - \frac{1}{n}\sum_i \frac{\sigma_i}{L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)}} \\
&+ \frac{1}{n}\sum_i \frac{-\sigma_i \int \frac{s}{1+sm_{1n,c}(L_+)}\mathrm{d}F(s) + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)}}{(L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})(L_+ - \sigma_i \int \frac{s}{1+sm_{1n,c}(L_+)}\mathrm{d}F(s))}
\end{aligned} \tag{S.7}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_i \frac{\sigma_i(L_+ - \widehat{L}_+)}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})(L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})} \\
&+ \frac{1}{n}\sum_i \frac{\sigma_i}{\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)}} - \frac{1}{n}\sum_i \frac{\sigma_i}{\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)}} \\
&+ \frac{1}{n}\sum_i \frac{\sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s) - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)}}{(L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})(L_+ - \sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s))} \\
&= \frac{1}{n}\sum_i \frac{\sigma_i(L_+ - \widehat{L}_+)}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})(L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})} \\
&+ \frac{1}{n}\sum_i \frac{-\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))(1+\xi_j^2 m_{1n,c}(L_+))}}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})} \\
&+ \frac{1}{n}\sum_i \frac{-\sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s) + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)}}{(L_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})(L_+ - \sigma_i \int \frac{s}{1+s m_{1n,c}(L_+)}\mathrm{d}F(s))}
\end{aligned}
$$

$$
:= \mathsf{T}_1 + \mathsf{T}_2 + \mathsf{T}_3. \tag{S.8}
$$

For the term T_1, by (S.6), Assumption S.1.1 and (S.27), we can see that

$$
\mathsf{T}_1 = \mathsf{C}_1(L_+ - \widehat{L}_+) + \mathrm{O}_{\mathbb{P}}(n^{-1}). \tag{S.9}
$$

For the term T_2, we see that

$$
\begin{aligned}
\mathsf{T}_2 &= \frac{1}{n}\sum_i \frac{-\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))(1+\xi_j^2 m_{1n,c}(L_+))}}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})^2} \\
&+ \frac{1}{n}\sum_i \frac{\left(\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))(1+\xi_j^2 m_{1n,c}(L_+))}\right)^2 (m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))^2}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})^2(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})} \\
&= \frac{1}{n}\sum_i \frac{-\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))^2}}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})^2} + \frac{1}{n}\sum_i \frac{-\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))^2}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))^2(1+\xi_j^2 m_{1n,c}(L_+))}}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})^2} \\
&+ \frac{1}{n}\sum_i \frac{\left(\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))(1+\xi_j^2 m_{1n,c}(L_+))}\right)^2 (m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+))^2}{(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(L_+)})^2} \\
&= -(m_{1n}(\widehat{L}_+) - m_{1n,c}(L_+)) + \mathrm{O}_{\mathbb{P}}(n^{-1}),
\end{aligned} \tag{S.10}
$$

where in the last step we used the second equation of (S.1) for the first term of (S.10), and (S.6), Theorem S.1.9, (S.27) and Assumption S.1.1 for the second and third terms. Similarly, for T_3, we have that

$$
\mathsf{T}_3 = \mathsf{C}_1\mathcal{X} + \mathrm{O}_{\mathbb{P}}(n^{-1}). \tag{S.11}
$$

Insert (S.9), (S.10) and (S.11) into (S.7), we can conclude the proof. □

Then we prove (S.6) to complete step two and the proof of the theorem.

Step two: To prove (S.6), we first rewrite (S.1) and (S.2) a little bit. Recall (S.3). We find that (S.1) can be rewritten as

$$F_n(m_{1n}(\widehat{L}_+), \widehat{L}_+) = 0, \quad \frac{\partial F_n}{\partial x}(m_{1n}(\widehat{L}_+), \widehat{L}_+) = 0,$$

where we denote

$$F_n(x,y) = \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-y + \frac{\sigma_i}{n}\sum_{j=1}^{n}\frac{\xi_j^2}{1+x\xi_j^2}} - x. \tag{S.12}$$

Similarly, (S.2) can be rewritten as

$$F_{n,c}(m_{1n,c}(L_+), L_+) = 0, \quad \frac{\partial F_{n,c}}{\partial x}(m_{1n,c}(L_+), L_+) = 0,$$

where we denote

$$F_{n,c}(x,y) = \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-y + \sigma_i \int \frac{s}{1+xs} dF(s)} - x.$$

For pair $(\widetilde{x}, \widetilde{y})$ so that $\widetilde{x} > -l^{-1}$ (recall (S.4)), as long as they satisfy Assumption S.1.1 in the sense that $\min_{1\leqslant i \leqslant p} |\tilde{y} - \sigma_i \int \frac{s}{1+\tilde{x}s} dF(s)| \geqslant \tau$, by (S.27), we find that

$$|F_{n,c}(\tilde{x},\tilde{y}) - F_n(\tilde{x},\tilde{y})| + \left|\frac{\partial F_{n,c}}{\partial x}(\tilde{x},\tilde{y}) - \frac{\partial F_n}{\partial x}(\tilde{x},\tilde{y})\right| + \left|\frac{\partial F_{n,c}}{\partial y}(\tilde{x},\tilde{y}) - \frac{\partial F_n}{\partial y}(\tilde{x},\tilde{y})\right| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}). \tag{S.13}$$

Set $(x_0, y_0) = (m_{1n,c}(L_+), L_+)$. Then we have that

$$F_{n,c}(x_0,y_0) = 0, \quad \frac{\partial F_{n,c}}{\partial x}(x_0,y_0) = 0, \quad 0 < \frac{\partial F_{n,c}}{\partial y}(x_0,y_0) < \infty, \quad \frac{\partial^2 F_n}{\partial y^2}(x_0,y_0) < 0. \tag{S.14}$$

It suffices to prove the following lemma.

LEMMA S.4.2. *There exists a pair* (x_1, y_1) *with condition* $|x_1 - x_0| + |y_1 - y_0| = \mathrm{O}_{\mathbb{P}}(n^{-1/2})$ *such that with probability* $1 - \mathrm{o}(1)$

$$F_n(x_1,y_1) = 0, \quad \frac{\partial F_n}{\partial x}(x_1,y_1) = 0. \tag{S.15}$$

With Lemma S.4.2, according to (S.1) and Theorem 6.2, we see that (S.6) holds. In the rest, we prove Lemma S.4.2 using (S.13).

Proof of Lemma S.4.2. For some small $\epsilon > 0$, we consider the probability event Ξ so that (S.27) holds and (S.13) reads as

$$|F_{n,c}(\tilde{x},\tilde{y}) - F_n(\tilde{x},\tilde{y})| + \left|\frac{\partial F_{n,c}}{\partial x}(\tilde{x},\tilde{y}) - \frac{\partial F_n}{\partial x}(\tilde{x},\tilde{y})\right| + \left|\frac{\partial F_{n,c}}{\partial y}(\tilde{x},\tilde{y}) - \frac{\partial F_n}{\partial y}(\tilde{x},\tilde{y})\right| = \mathrm{O}(n^{-1/2+\epsilon}). \tag{S.16}$$

We have seen that $\mathbb{P}(\Xi) = 1 - \mathrm{o}(1)$. Now we fix a realization $\{\xi_i^2\} \in \Xi$ so that the discussions below are purely deterministic.

For the above fixed constant $\epsilon > 0$, we set the region

$$\mathcal{N}(x,y) := \{(x,y) : |x - x_0| + |y - y_0| \leqslant n^{-1/2+\epsilon}\},$$

To prove the first part of (S.15), it suffices to prove that there exists a solution of $F_n(x,y) = 0$ in the region $\mathcal{N}(x,y)$. By Bolzano's theorem, we see that for sufficiently large n, we can find two points (x_{11}, y_{11}) and (x_{12}, y_{12}) on $\mathcal{N}(x,y)$ so that $F_{n,c}(x_{11}, y_{11}) < 0,\ F_{n,c}(x_{12}, y_{12}) > 0$. Together with (S.16), we see that $F_n(x_{11}, y_{11}) < 0,\ F_n(x_{12}, y_{12}) > 0$. Therefore, by continuity, we can find some point (x', y') so that $F_n(x', y') = 0$. Repeating the above procedure, by implicit function theorem, we find that there exists a curve $x \equiv x(y)$ on $\mathcal{N}(x,y)$ so that $F_n(x,y) = 0$. Similarly, we can show that there exists another curve $\widehat{x} \equiv \widehat{x}(\widehat{y})$ on $\mathcal{N}(\widehat{x}, \widehat{y})$ so that the second part of (S.15) holds in the sense that $\partial F_n(\widehat{x}, \widehat{y})/\partial \widehat{x} = 0$.

In order to show (S.15), we need to prove that the curves (x,y) and $(\widehat{x}, \widehat{y})$ must have at least one intersection in the region $\mathcal{N}(x,y)$. We prove by contradiction. Otherwise, the curve (x,y) will lie in one of the areas separated by $(\widehat{x}, \widehat{y})$ with strictly $\partial F_n(x,y)/\partial x < 0$ or $\partial F_n(x,y)/\partial x > 0$. By (S.14), we see that $\mathcal{N}(x,y),\ \partial F_n(x,y)/\partial y > 0$. Without loss of generality, we assume $\partial F_n(x,y)/\partial x < 0$. On the one hand, as $F_n(x,y) = 0$, one may conclude that for small neighbor around the points on (x,y), it holds that $\mathrm{d}x/\mathrm{d}y > 0$. On the other hand, taking the derivative $F_n(x,y)$ with respect to y, we obtain that

$$\frac{\mathrm{d}x}{\mathrm{d}y} \times \Big(\frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{(1+x\xi_j^2)^2}}{(-y + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+x\xi_j^2})^2} - 1\Big) = 0,$$

which implies $\partial F_n(x,y)/\partial x = 0$ and gives the contradiction. This concludes our proof. □

S.4.3. Proof of the results in Section 3.3

Proof of Corollary 3.6. As calculated by [53, Lemma 1], we have

$$\mathbb{P}\big(|\xi^2 - 1| > 2\sqrt{\frac{s}{T}} + 2\frac{s}{T}\big) \le e^{-s}.$$

Setting $s = \log^\tau n$ for some large constant $\tau > 0$, we observe that $\xi^2 \in (1 - 2\log^{\tau/2} nT^{-1/2}, 1 + 2\log^{\tau/2} nT^{-1/2})$ with high probability. Then, it is free to do truncation of ξ^2 on the support $(1 - 2\log^{\tau/2} nT^{-1/2}, 1 + 2\log^{\tau/2} nT^{-1/2})$ with negligible error after choosing τ large. Due to this reason, we regard ξ^2 has bounded support on $(1 - 2\log^{\tau/2} nT^{-1/2}, 1 + 2\log^{\tau/2} nT^{-1/2})$ in the following. Then a direct calculation shows that

$$\begin{aligned}
\mathsf{v} &= \int_{1-2\log^{\tau/2} nT^{-1/2}}^{1+2\log^{\tau/2} nT^{-1/2}} \Big(\frac{s}{1 + s m_{1n,c}(L_+)}\Big)^2 \mathrm{d}F(s) - \Big(\int_{1-2\log^{\tau/2} nT^{-1/2}}^{1+2\log^{\tau/2} nT^{-1/2}} \frac{s}{1 + s m_{1n,c}(L_+)} \mathrm{d}F(s)\Big)^2 \\
&\lesssim (1 + 2\log^{\tau/2} nT^{-1/2})^2 - (1 - 2\log^{\tau/2} nT^{-1/2})^2 \\
&\lesssim \log^{\tau/2} nT^{-1/2}.
\end{aligned}$$

If we set T as some proportion of size n, then it holds that $\mathsf{v} = \mathrm{o}(n^{-1/3})$. On the other hand, as ξ^2 concentrates tightly around its mean, the system equations (6.2) for each bootstrapped matrices Q_b will reduce to

$$\frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i}{-z + \frac{\sigma_i}{1+m_{1n}(z)} + \mathrm{O}(\log^{\tau/2} nT^{-1/2})} - m_{1n}(z) = 0.$$

Comparing this reduced equation with the one in [26], we find that $m_{1n}(z)$ establish the similar system equation as covariance matrix S asymptotically. As a consequence, $n^{2/3}(\lambda_{b,1} -$

$\widehat{L}_+$) inherits the typical TW limit from the top eigenvalue of S. Then, a natural candidate to estimate $\widehat{L}_+$ is the sample mean of $\{\lambda_{b,1}\}_{1\le b\le B}$. We write $\widetilde{L}_+ := B^{-1}\sum_{i=1}^{B}\lambda_{b,i}$. By strong law of large number, it is easy to find that

$$\widetilde{L}_+ = \widehat{L}_+ + \mathrm{O}_p(B^{-1/2}).$$

Since we have set that $B = n^{5/3}$, it holds that $n^{2/3}\times \mathrm{O}_p(B^{-1/2}) = \mathrm{o}_p(1)$. Using the Slutsky's theorem, together with asymptotic TW distribution for $n^{2/3}(\lambda_{b,1} - \widehat{L}_+)$, we have $n^{2/3}(\lambda_{b,1} - \widetilde{L}_+)$ has the same TW limit in distribution.

For the unconditional result, as illustrated above, the difference of $|\widehat{L}_+ - L_+|$ is much smaller than $\mathrm{O}(n^{-2/3})$ since v decays sufficiently fast due to our choice of T. Therefore, the limiting spectral distribution of each Q_b is essentially dominated by the TW distribution. We omit further details here.

□

S.4.4. Testing for the non-spiked structure

In this section, we use Corollary 3.6 and Algorithm 1 to test non-spiked structure of a given covariance matrix. This problem is particularly challenging if the data entails the weak spikes (comparing to the situation of strong spikes in Section 4). The insights to apply the above results to generate a feasible algorithm to detect weak spikes lie in that if the data matrix S exhibits non-spiked structure, its leading eigenvalues are expected to follow the Tracy-Widom (TW) distribution, with spacings exhibiting fluctuations at the TW scale, $n^{-2/3}$. This insight motivates the use of Algorithm 1 to bootstrap the top r eigenvalues of S for some pre-given integer r. By leveraging the empirical distributions of r bootstrapped TW eigenvalues obtained from Algorithm 1, we can construct r confidence intervals. Statistical inference is then performed by testing whether there exists eigenvalue of S lie outside the nearest bootstrapped confidence intervals, thereby indicating potential rejection of the null hypothesis.

To be specific, we state a covariance matrix exhibits possibly weak spiked structure if the leading eigenvalues distract from the bulk eigenvalues beyond the level of Tracy-Widom fluctuation but still remain a small or bounded region. Within this section, we consider the sample covariance matrix $\widehat{S} = \widehat{\Sigma}^{1/2}XX^*\widehat{\Sigma}^{1/2}$, where $\widehat{\Sigma}$ has the spectral decomposition

$$\widehat{\Sigma} = \sum_{j=1}^{p}\widehat{\sigma}_j \mathbf{v}_j\mathbf{v}_j^*. \tag{S.17}$$

We suppose that for fixed constant $r\ge 0$, the r largest eigenvalues $\widehat{\sigma}_1 \ge \widehat{\sigma}_2 \ge \cdots \ge \widehat{\sigma}_r$, may exhibit a significant separation from the remaining eigenvalues $\widehat{\sigma}_{r+1}\ge\cdots\ge\widehat{\sigma}_p$, which are clustered within a dense interval. To be specific, we assume that for some constant $\tau > 1$ and a large value $s > 0$ (bounded or possibly divergent with some rates of n),

$$\widehat{\sigma}_r - \widehat{\sigma}_{r+1} > s, \quad \tau \ge \widehat{\sigma}_{r+1} \ge \widehat{\sigma}_p \ge \tau^{-1}.$$

We consider the hypothesis testing problem

$$\mathbf{H}_0 : r = 0 \quad \text{vs} \quad \mathbf{H}_a : r \geqslant 1. \tag{S.18}$$

Based on Algorithm 1 and Corollary 3.6, we can propose the following algorithm to test (S.18).

Algorithm 4 Resampling testing for (S.18)

Inputs: Given integer $r_0 > 0$, original data $\widehat{S}$, type I error α.
Step One: Plug in $\widehat{S}$ and run Algorithm 1 to bootstrap first leading r_0 eigenvalues from $\widehat{S}$. Suppose for each $1 \leq i \leq r_0$, we obtain the sequence bootstrapped eigenvalues $\{\lambda_{b,i}\}_{1 \leq b \leq B}$. Calculate $F_{\mathsf{TW}}^{(i)}(x) := B^{-1}\#\{b : n^{2/3}(\lambda_{b,i} - \widetilde{L}_+^{(i)}) \leq x\}$ as in Algorithm 1, where $\widetilde{L}_+^{(i)} = B^{-1}\sum_{b=1}^{B} \lambda_{b,i}$. For given type I error α, calculate the $1-\alpha$ quantile of each $F_{\mathsf{TW}}^{(i)}(x)$, denoted as $\{x_{i,1-\alpha}\}_{1 \leq i \leq r_0}$. Construct the intervals $\mathcal{I}_i := [\widetilde{L}_+^{(i)} - n^{-2/3}x_{i,1-\alpha/2}, \widetilde{L}_+^{(i)} + n^{-2/3}x_{i,1-\alpha/2}]$ for each $1 \leq i \leq r_0$.
Step Two: Calculate the eigenvalues of the original covariance matrix $\widehat{S}$ and order them in $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_n \geq 0$.
Step Three: Set $a = r_0$, run the following iteration:

if $\widehat{\lambda}_{a-1} \in \mathcal{I}_a$ **then**
  Update $a = a - 1$ and repeat till.
if There is no eigenvalue $\widehat{\lambda}_i, i < a$ in the interval $\mathcal{I}_a$ **then**
  Record $r = a - 1$.
else if $a = 0$ **then**
  Record $r = 0$.
end

Output: Reject $\mathbf{H}_0$ in (S.18) if $r > 0$.

COROLLARY S.4.3. *Suppose the assumptions of Theorem 3.3 and Corollary 3.6 hold. Under $\mathbf{H}_0$, we have*

$$\lim_{n\to\infty} \mathbb{P}(r = 0) = 1.$$

On the other hand, under $\mathbf{H}_a$, we have

$$\lim_{n\to\infty} \mathbb{P}(r > 0) = 1.$$

PROOF. The proof mainly bases on Corollary 3.6 for non-spiked eigenvalues. We omit further details here. □

□

APPENDIX S.5: PROOF OF THE RESULTS OF SECTION 4

In this section, we prove the results of Section 4.

S.5.1. Proof of the results in Section 4.1

Proof of Theorem 4.1. Recall the bootstrapped sample covariance matrix $\widetilde{Q}$, denote $\widetilde{Y} = \widetilde{\Sigma}^{1/2}XD$. Then we write $\widetilde{Q} := \widetilde{Y}\widetilde{Y}^*$ and $\widetilde{\mathcal{Q}} := \widetilde{Y}^*\widetilde{Y}$. Since these two matrices have the same non-zero eigenvalues, we focus on the later one for convenience. For the spiked covariance matrix model in (2.12), we decompose it as follows

$$\widetilde{\Sigma} := \Sigma_s + \Sigma_o,$$

where we denote the two $p\times p$ matrices as

$$\Sigma_s := \sum_{i=1}^r \tilde{\sigma}_i \mathbf{v}_i \mathbf{v}_i^* \equiv V_1 \Lambda_s V_1^*, \quad \Sigma_o := \sum_{i=r+1}^p \sigma_i \mathbf{v}_i \mathbf{v}_i^* \equiv V_2 \Lambda_o V_2^*. \tag{S.1}$$

Consequently, we can decompose $\widetilde{\mathcal{Q}}$ as follows

$$\widetilde{\mathcal{Q}} = DX^*\tilde{\Sigma}XD = DX^*\Sigma_s XD + DX^*\Sigma_o XD.$$

Note that with high probability

$$\|DX^*\Sigma_o XD\| = \|D^2X^*\Sigma_0 X\| \leqslant \|D^2\|\|X^*\Sigma_0 X\| \leq \sigma_r \xi_{(1)}^2 \|X^*X\| \sim \xi_{(1)}^2,$$

where in the last step we used [70] that $\|X^*X\|$ is bounded from above with high probability. Using (S.25) and (S.26) as well as Weyl's inequality, we see that from the assumption of (4.4) that, for $1\leqslant i\leqslant r$,

$$\frac{\mu_i - \lambda_i(DX^*\Sigma_s XD)}{\widetilde{\sigma}_i} = \mathrm{o}_{\mathbb{P}}(1). \tag{S.2}$$

Then we consider the first few largest eigenvalues of $DX^*\Sigma_s XD$, or equivalently those of $\Sigma_s^{1/2}XD^2X^*\Sigma_s^{1/2}$. By a discussion similar to Lemma D.1 of [27], we find that if λ is an eigenvalues of $\Sigma_s^{1/2}XD^2X^*\Sigma_s^{1/2}$, recalling (S.1), we have that

$$\det(V_1^*XD^2X^*V_1 - \lambda\Lambda_s^{-1}) = 0. \tag{S.3}$$

Note that

$$V_1^*XD^2X^*V_1 - \mathbb{E}\xi^2 I_r = V_1^*X(D^2-\mathbb{E}\xi^2 I_r)X^*V_1 + \mathbb{E}\xi^2\times(V_1^*XX^*V_1 - I_r),$$

where I_r is a $r\times r$ identity matrix. For the first term in decomposition, we observe that

$$X(D^2-\mathbb{E}\xi^2 I_r)X^* = \begin{pmatrix} \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i1}^2 & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i1}x_{i2} & \cdots & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i1}x_{ir} \\ \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i2}x_{i1} & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i2}^2 & \cdots & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{i2}x_{ir} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{ir}x_{i1} & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{ir}x_{i2} & \cdots & \sum_i(\xi_i^2-\mathbb{E}\xi^2)x_{ir}^2 \end{pmatrix}_{r\times r},$$

By Assumptions 2.1 and 2.2, and notice that r is finite, a straightforward calculations indicate that

$$\|X(D^2-\mathbb{E}\xi^2 I_r)X^*\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}).$$

It follows that

$$\|V_1^*X(D^2-\mathbb{E}\xi^2 I_r)X^*V_1\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}).$$

On the other hand, by Theorem 7.1 of [7], one has

$$\|V_1^*XX^*V_1 - I_r\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}).$$

As a consequence, we conclude that

$$\|V_1^*XD^2X^*V_1 - \mathbb{E}\xi^2 I_r\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}),$$

Together with (S.3), we conclude that for $1\leqslant i\leqslant r$

$$\frac{\lambda_i(DX^*\Sigma_s XD)}{\widetilde{\sigma}_i} = \mathbb{E}\xi^2 + \mathrm{o}_{\mathbb{P}}(1). \tag{S.4}$$

Combining (S.2), we obtain the first order result for the limiting property of $\mu_i/\widetilde{\sigma}_i$.

Then we proceed with the second order results for μ_i. The proof follows from strategies similar to Theorem 3.3 of [74]. We focus on explaining the main ideas and omit the details. The core of the proof is to introduce the auxiliary quantities $\theta_i, 1 \leqslant i \leqslant r$, where for each $1 \leqslant i \leqslant r$, θ_i satisfies the equation

$$\frac{\theta_i}{\widetilde{\sigma}_i} = \Big(1 - \frac{1}{n\theta_i} \sum_{j=r+1}^{p} \frac{\sigma_j}{1 - \widetilde{\sigma}_i^{-1}\sigma_j}\Big)^{-1}. \tag{S.5}$$

With the restriction that $\theta_i \in [\widetilde{\sigma}_i, 2\widetilde{\sigma}_i]$, the existence of uniqueness of θ_i have been justified in [15, 74]. Furthermore, under the assumption of (4.4), we can conclude from equation (2.10) of [15] that

$$\theta_i/\widetilde{\sigma}_i = 1 + \mathrm{o}(1), 1 \leqslant i \leqslant r. \tag{S.6}$$

In order to establish the asymptotics of μ_i/θ_i, for notational convenience, we now work with the rescaled matrix

$$\check{\mathcal{Q}} := \check{D}X^*\widetilde{\Sigma}X\check{D}, \ \check{D}^2 := (\mathbb{E}\xi^2)^{-1}D^2,$$

whose eigenvalues are denoted as $\check{\lambda}_1 \geqslant \check{\lambda}_2 \geqslant \cdots \geqslant \check{\lambda}_{\{p\wedge n\}} > 0$. Note that $\mu_i = \mathbb{E}\xi^2\check{\lambda}_i$. By a discussion similar to (S.3) and (S.12), using (S.1), we find that $\check{\lambda}_i, 1 \leqslant i \leqslant r$ satisfy the equation

$$\det(\Lambda_s^{-1} - V_1^* X\check{D}(\check{\lambda}_i I - \check{D}X^*\Sigma_o X\check{D})^{-1}\check{D}X^*V_1) = 0.$$

Denote $\mathbf{B}(x) := xI - \check{D}X^*\Sigma_o X\check{D}$ and $\delta_i = (\check{\lambda}_i - \theta_i)/\theta_i$, the above determinant can be rewritten into

$$\det(\theta_i\Lambda_s^{-1} - \theta_i V_1^* X\check{D}\mathbf{B}^{-1}(\theta_i)\check{D}X^*V_1 + \delta_i\theta_i^2 V_1^*X\check{D}\mathbf{B}^{-1}(\check{\lambda}_i)\mathbf{B}^{-1}(\theta_i)\check{D}X^*V_1) = 0. \tag{S.7}$$

Following the procedure in Section 7.1 of [15] or Lemma C.5 of [74], we find that for $1 \leqslant i, l \leqslant r$ (recall that $\widetilde{\Sigma}$ is assumed to be diagonal)

$$\theta_i \mathbf{e}_i^* V_1^* X\check{D}\mathbf{B}^{-1}(\theta_i)\check{D}X^*V_1\mathbf{e}_l = \mathbf{1}(l=i)\big(\sum_{j=1}^{n} x_{kj}^2\xi_j^2/\mathbb{E}\xi^2 - \frac{1}{n}\sum_{j=1}^{n}\xi_j^2/\mathbb{E}\xi^2 + \zeta_i\big) + \mathrm{o}_{\mathbb{P}}(n^{-1/2}),$$

where ζ_i is a random quantity associated with θ_i satisfying

$$\zeta_i := \frac{1}{n}\sum_{j=1}^{n}\frac{\xi_j^2}{\mathbb{E}\xi^2}\Big(1 - \frac{\xi_j^2}{n\theta_i\mathbb{E}\xi^2}\sum_{k=r+1}^{p}\frac{\sigma_k}{1-\theta_i^{-1}\sigma_k\zeta_i}\Big)^{-1}. \tag{S.8}$$

Similarly, by a discussion similar to Lemma C.6 of [74], we conclude that

$$\delta_i\theta_i^2[V_1^*X\check{D}\mathbf{B}^{-1}(\check{\lambda}_i)\mathbf{B}^{-1}(\theta_i)\check{D}X^*V_1]_{il} = \delta_i \times (\mathbf{1}(l=i) + \mathrm{o}_{\mathbb{P}}(1)).$$

Inserting the above two controls into (S.7), by the assumption of (2.13), using Leibniz's formula for determinant, one has that

$$\delta_i(1+\mathrm{o}_{\mathbb{P}}(1)) = (\mathbb{E}\xi^2)^{-1}\mathbf{v}_i^* X D^2 X^* \mathbf{v}_i - \frac{1}{n}\sum_{j=1}^{n}\xi_j^2/\mathbb{E}\xi^2 - \frac{\theta_i}{\tilde{\sigma}_i} + \zeta_i + \mathrm{o}_{\mathbb{P}}(n^{-1/2}). \tag{S.9}$$

By a similar argument as in Lemma 3.2 of [74] and notice that $\xi_j^2 \ll \widetilde{\sigma}_i, 1 \le j \le n$, one has

$$\begin{aligned}\zeta_i - \frac{\theta_i}{\widetilde{\sigma}_i} &= \frac{1}{n}\sum_{j=1}^{n} \xi_j^2/\mathbb{E}\xi^2 - 1 + \Big(\frac{1}{n}\sum_{k=r+1}^{p} \frac{\sigma_k}{\widetilde{\sigma}_i} \times \mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^2\Big) \times (1+\mathrm{o}(1)) \\ &+ \Big(\big(\frac{1}{n}\sum_{k=r+1}^{p} \frac{\sigma_k}{\widetilde{\sigma}_i}\big)^2 \times \mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^3\Big) \times (1+\mathrm{o}(1)) \\ &+ \big(\frac{1}{n}\sum_{k=r+1}^{p} \frac{\sigma_k}{\widetilde{\sigma}_i}\big)^2 \times \mathrm{o}_{\mathbb{P}}(n^{-1/2}) + \mathrm{o}_{\mathbb{P}}(n^{-1/2}).\end{aligned} \tag{S.10}$$

Combining the above results, we conclude that,

$$\frac{\check{\lambda}_i - \theta_i}{\theta_i} = (\mathbb{E}\xi^2)^{-1}\mathbf{v}_i^* X D^2 X^* \mathbf{v}_i - 1 + \delta_{\mathrm{c}}/\mathbb{E}\xi^2 + \mathrm{o}_{\mathbb{P}}(n^{-1/2}),$$

where

$$\delta_{\mathrm{c}} := \mathbb{E}\xi^2 \times \big(\frac{1}{n}\sum_{k=r+1}^{p} \frac{\sigma_k}{\widetilde{\sigma}_i} \times \mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^2\big) \times (1+\mathrm{o}(1)) + \mathbb{E}\xi^2 \times \Big(\big(\frac{1}{n}\sum_{k=r+1}^{p} \frac{\sigma_k}{\widetilde{\sigma}_i}\big)^2 \times \mathbb{E}(\xi_1^2/\mathbb{E}\xi^2-1)^3\Big) \times (1+\mathrm{o}(1)),$$

is defined as a deterministic correction quantity. By Assumption 2.1 and using central limit theorem, we can conclude the proof of (4.5).

Finally, we briefly illustrate the proof of (4.6). The argument follows closely from a discussion similar to the proof of [27, Theorem 3.7], or [50, Theorem 2.7], or [14, Theorem 2.7], or [23, Theorem 3.6]. Due to similarity, we only sketch the proof strategies, provide the key ingredients and point out the main differences. In fact, our proof will be easier since the spikes are much larger than the edges and we only consider the first few extremal non-outlier eigenvalues. As discussed in [27, Appendix D], or [50, Section 6], or [14, Section 4], the proof consists of the following three steps.

(i). We first find the permissible regions in which contain the eigenvalues of $\widetilde{Q}$ with high probability.
(ii). Then we apply a counting argument to a special case (where all the spikes are well-separate), and show that the results hold under this special case.
(iii). Finally we use a continuity argument to extend the results in (ii) to the general case using the gaps in the permissible regions.

In what follows, we choose a realization $\{\xi_i^2\} \in \Omega$ so that (S.27) holds with $1-\mathrm{o}(1)$ probability as in Lemma S.1.12. With this restriction, m_{1n} and ϑ_1 in (6.5) are purely deterministic. Recall d_1 in (S.4) and the ϵ used therein. Due to similarity, we focus on the polynomial decay setting (2.3). The exponential decay case can be handled similarly.

For Step (i), to find the permissible region, for some large constant $\mathsf{C} > 0$, we denote the set for $1 \leqslant i \leqslant k$

$$\Gamma_i := \Big\{x \in [\lambda_i, \vartheta_1 + n^{-1/2+2\epsilon} d_1] : \ \mathrm{dist}(x, \mathrm{spec}(Q)) > \mathsf{C} n^{-1/2+2\epsilon} d_1\Big\}, \tag{S.11}$$

where spec(Q) stands for the spectrum of Q. The results of Step (i) can be summarized as follows.

LEMMA S.5.1. *There exists some constant $\mathsf{C} > 0$ so that the set $\cup_i \Gamma_i$ contains no eigenvalue of $\widetilde{Q}$.*

PROOF. The proof is similar to that of Lemma D.4 of [27] or Lemma 5.4 of [23] and we only sketch the key points here. Notice that $\widetilde{\Sigma}$ has the same size as Σ, we can decompose that $\widetilde{\Sigma} = \Sigma_1 + \Sigma$ with $\Sigma_1 = \widetilde{\Sigma} - \Sigma = U_1 \Lambda_1 U_1^*$, where we recall that $\widetilde{\Sigma}$ is constructed based on Σ. Here Λ_1 is an $r \times r$ matrix containing the nonzero eigenvalues of Σ_1 and U_1 is the $p \times r$ matrix containing the first r associated eigenvectors in $\mathbb{R}^p$. Under the assumption of (4.4) and (2.11), we see that Λ_1 is invertible when n is sufficiently large. By a discussion similar to (S.3), we see that x is an eigenvalue of $\widetilde{\mathcal{Q}}$ but not $\mathcal{Q}$ if and only if

$$\det(I - \Lambda_1^{1/2} U_1^* X D (xI - DX^* \Sigma X D)^{-1} D X^* U_1 \Lambda_1^{1/2}) = 0. \tag{S.12}$$

Moreover, for $x \in \Gamma_i, 1 \leqslant i \leqslant k$ and $\eta := n^{-1/2}$, we define $z_x = x + \mathrm{i}\eta$. According to Proposition S.2.1, we have that with high probability

$$\left\| (DX^* \Sigma X D - z_x I)^{-1} + z^{-1}(I + m_{1n}(z_x) D^2)^{-1} \right\|_\infty = \mathrm{O}(n^{-1/2 - 1/\alpha + \epsilon}).$$

On the other hand, we observe that

$$U_1^* X D^2 X^* U_1 = \mathbb{E}\xi^2 I_r + \mathrm{o}_{\mathbb{P}}(1).$$

Using the fact r is finite and the definition of $m_{2n}(z)$ in (6.2), we find that

$$\left\| z^{-1} U_1^* X D (I + m_{1n}(z_x) D^2)^{-1} D X^* U_1 - m_{2n}(z_x) I \right\|_\infty = \mathrm{O}_{\mathbb{P}}(n^{-1/2 - 1/\alpha + 2\epsilon}),$$

where we used the fact that $|z_x| \asymp \vartheta_1$ and (S.6). According to a discussion similar to equation (D.29) of [27] and Lemma S.1.2, we see that for $t = 1, 2$,

$$m_{tn}(z_x) - m_{tn}(x) \asymp \operatorname{Im} m_{tn}(z) = \mathrm{O}(n^{-1/2 - 1/\alpha + \epsilon}).$$

Combining all the above controls, we find that for some constant $C > 0$

$$\left\| I - \Lambda_1^{1/2} U_1^* X D (xI - DX^* \Sigma X D)^{-1} D X^* U_1 \Lambda_1^{1/2} \right\|_\infty = C \max_{1 \leqslant j \leqslant r} |m_{2n}(z_x) + \widetilde{\sigma}_j^{-1}| + \mathrm{O}_{\mathbb{P}}(n^{-1/2 - 1/\alpha + \epsilon}).$$

Since $x \in \Gamma_i$, together with Lemma S.1.2, we see that $|m_{2n}(z_x) + \widetilde{\sigma}_j^{-1}| \gg n^{-1/2 - 1/\alpha + \epsilon}$. This implies that x is not an eigenvalue of $\widetilde{\mathcal{Q}}$ and completes the proof. □

As mentioned in the proof of Theorem 2.7 of [14], once Step (i) is done, Steps (ii) and (iii) are more standard. For Step (ii), together with the interlacing results as in Lemma C.3 of [27], we perform the counting argument to prove (4.6) for a special case assuming $\widetilde{\sigma}_1 > \widetilde{\sigma}_2 > \cdots > \widetilde{\sigma}_r$. The details can be found in Lemma D.5 of [27] or Lemma 5.5 of [23]. For Step (iii), we use a continuity argument for all possible configurations $\{\widetilde{\sigma}_i\}_{1 \leqslant i \leqslant r}$. The details can be found in the proof of Theorem 3.7 of [27]. Since most of the arguments can be made verbatim following lines of the counterparts of [27] or [23] or [14] or [50], we omit further details. This completes the proof.

□

Proof of Theorem 4.3. The proof follows from strategies similar to Theorem 3.5 of [74]. We focus on explaining the main ideas and omit the details. Before we proceed to the main step, we need the following results for the limiting behavior of spiked eigenvalues in $\widetilde{S}$.

LEMMA S.5.2. *Under assumptions in Theorem 4.1 with $\widetilde{\sigma}_r \gg \mathsf{T}$. We have for $1 \leq i \leq r$, $\widehat{\mu}_i$ is closed to $\widetilde{\sigma}_i$ in the sense that*

$$\frac{\widehat{\mu}_i}{\widetilde{\sigma}_i} = 1 + \mathrm{o}_{\mathbb{P}}(1),$$

while $\widehat{\mu}_{r+1} = \mathrm{O}_{\mathbb{P}}(1)$. *Moreover,* $\widehat{\mu}_i$ *admits the limiting representation with* θ_i *as*

$$\frac{\widehat{\mu}_i}{\theta_i} - 1 = \mathbf{v}_i^* X X^* \mathbf{v}_i - 1 + \mathrm{o}_{\mathbb{P}}(1),$$

for $1 \leq i \leq r$.

PROOF. Notice that for $1 \leqslant i \leqslant r$,

$$\frac{\widehat{\mu}_i}{\theta_i} = 1 + \mathbf{k}_i^* (V_1^* X X^* V_1 - I_r) \mathbf{k}_i + \mathrm{o}_{\mathbb{P}}(\frac{1}{\sqrt{n}}),$$

here $\mathbf{k}_i, 1 \leqslant i \leqslant r$, are the standard basis in $\mathbb{R}^r$. Then, the proof of Lemma S.5.2 is exactly the same as the one of Lemma 3.4 of [74] or Theorem 2.1 of [15]. We omit further details for simplicity. □

The core inputs of the proof of Theorem 4.3 are the results in Theorem 4.1 and Lemma S.5.2,

$$\frac{\check{\lambda}_i}{\theta_i} - 1 = (\mathbb{E}\xi^2)^{-1} \mathbf{v}_i^* X D^2 X^* \mathbf{v}_i - 1 + \delta_{\mathrm{c}}/\mathbb{E}\xi^2 + \mathrm{o}_{\mathbb{P}}(n^{-1/2})$$

$$\frac{\widehat{\mu}_i}{\theta_i} - 1 = \mathbf{v}_1^* X X^* \mathbf{v}_i - 1 + \mathrm{o}_{\mathbb{P}}(n^{-1/2}).$$

Then, we have

$$\begin{aligned}
\frac{\check{\lambda}_i}{\widehat{\mu}_i} = \frac{\check{\lambda}_i}{\theta_i} \times \frac{\theta_i}{\widehat{\mu}_i} &= \frac{1 + (\mathbb{E}\xi^2)^{-1} \mathbf{v}_i^* X D^2 X^* \mathbf{v}_i - 1 + \delta_{\mathrm{c}}/\mathbb{E}\xi^2 + \mathrm{o}_{\mathbb{P}}(n^{-1/2})}{1 + \mathbf{v}_1^* X X^* \mathbf{v}_i - 1 + \mathrm{o}_{\mathbb{P}}(n^{-1/2})} \\
&= \frac{(\mathbb{E}\xi^2)^{-1} \mathbf{v}_i^* X (D^2 - I) X^* \mathbf{v}_i + \delta_{\mathrm{c}}/\mathbb{E}\xi^2}{\mathbf{v}_1^* X X^* \mathbf{v}_i + \mathrm{o}_{\mathbb{P}}(n^{-1/2})} + (\mathbb{E}\xi^2)^{-1} + \mathrm{o}_{\mathbb{P}}(n^{-1/2}) \\
&= (\mathbb{E}\xi^2)^{-1} \mathbf{v}_i^* X (D^2 - I) X^* \mathbf{v}_i + \delta_{\mathrm{c}}/\mathbb{E}\xi^2 + (\mathbb{E}\xi^2)^{-1} + \mathrm{o}_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

It remains to find the limiting distribution of $\mathbf{v}_i^* X (D^2 - I) X^* \mathbf{v}_i$, which is a standard argument as in the proof of Theorem 3.5 in [74], using Assumptions 2.1 and 2.2. We omit further details here.

□

S.5.2. Proof of the results in Section 4.2

Proof of Corollary 4.5. From Algorithm 2, one may easily find from the strong law of large number that for each $1 \leq i \leq r$

$$\widehat{\mathsf{M}}_i = \mathsf{M}_i + \mathrm{O}_{\mathbb{P}}(B^{-1/2}), \quad \widehat{\mathsf{V}}_i = \mathsf{V}_i + \mathrm{O}_{\mathbb{P}}(B^{-1/2}).$$

Then this corollary can be concluded by Theorem 4.3, the Slutsky's theorem and the choice of B. □

APPENDIX S.6: PROOF OF SOME AUXILIARY LEMMAS

S.6.1. Preliminary estimates: Proof of Lemmas S.1.2 and S.1.5

S.6.1.1. Proof of Lemma S.1.2

Due to similarity, we only prove the results for the separable covariance i.i.d. data model when ξ^2 decays polynomially, i.e., when (S.3) and (2.3) hold. The other cases can be proved analogously and we omit the details.

Proof. We start with the first statement. We now abbreviate $F_n(m_{1n}(z)) \equiv F_n(m_{1n}(z), z)$ throughout the proof. For the real part, it suffices to prove that with high probability for some $0 < C_2 < 1 < C_1$

$$\operatorname{Re} m_{1n}(z) \in \left[-C_1 \frac{\phi\bar{\sigma}_1 E}{E^2+\eta^2}, -C_2 \frac{\phi\bar{\sigma}_1 E}{E^2+\eta^2}\right]. \tag{S.1}$$

Moreover, by continuity and Theorem 6.2, it suffices to prove the following inequalities

$$\operatorname{Re} F_n(-C_2\phi\bar{\sigma}_1 E(E^2+\eta^2)^{-1} + \mathrm{i}\operatorname{Im} m_{1n}(z)) < 0,\ \operatorname{Re} F_n(-C_1\phi\bar{\sigma}_1 E(E^2+\eta^2)^{-1} + \mathrm{i}\operatorname{Im} m_{1n}(z)) > 0. \tag{S.2}$$

We only focus on the first part. By definition, we have that

$$\begin{aligned}\operatorname{Re} F_n(m_{1n}(z)) = & -\operatorname{Re} m_{1n}(z) \\ & - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i \operatorname{Re}(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}(z)\xi_j^2})}{\operatorname{Re}^2(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}(z)\xi_j^2}) + \operatorname{Im}^2(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}(z)\xi_j^2})}.\end{aligned} \tag{S.3}$$

Note that

$$\operatorname{Re}\Big(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2}\Big) = E - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2(1+\xi_j^2\operatorname{Re} m_{1n})}{(1+\xi_j^2\operatorname{Re} m_{1n})^2 + \xi_j^4\operatorname{Im}^2 m_{1n}},$$

$$\operatorname{Im}\Big(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2}\Big) = \eta + \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^4\operatorname{Im} m_{1n}}{(1+\xi_j^2\operatorname{Re} m_{1n})^2 + \xi_j^4\operatorname{Im}^2 m_{1n}}.$$

By a discussion similar to (S.7), if $\operatorname{Re} m_{1n} = -C_2(\phi\bar{\sigma}_1 E)/(E^2+\eta^2)$, we have that

$$\begin{aligned}\operatorname{Re}\Big(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2}\Big) &\geq E(1-\mathrm{o}(1)),\\ \operatorname{Im}\Big(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2}\Big) &\leq \eta + E\times \mathrm{o}(1).\end{aligned} \tag{S.4}$$

Therefore, together with (S.3), we see that

$$\begin{aligned}\operatorname{Re} F_n(-C_2\phi\bar{\sigma}_1 E(E^2+\eta^2)^{-1} + \mathrm{i}\operatorname{Im} m_{1n}(z)) &\leq C_2\phi\bar{\sigma}_1 \frac{E}{(E^2+\eta^2)} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i \times E(1-\mathrm{o}(1))}{E^2+(\eta+E\times\mathrm{o}(1))^2}\\ &\leq (C_2 - 1 + \mathrm{o}(1))\frac{\phi\bar{\sigma}_1 E}{(E^2+\eta^2)} < 0,\end{aligned} \tag{S.5}$$

for sufficient large n. This completes the discussion for the real part. For the complex part, the idea is similar and it suffices to prove that when $z \in \mathbf{D}_u$, for some constants $C_1, C_2 > 0$

$$\operatorname{Im} m_{1n}(z) \in \left[C_1 \frac{\eta\phi\bar{\sigma}_1}{E^2+\eta^2}, C_2\eta\,|\operatorname{Re} m_{1n}(z)|\right]. \tag{S.6}$$

Equivalently, it suffices to prove that

$$\operatorname{Im} F_n(\operatorname{Re} m_{1n}(z) + \mathrm{i}C_2\eta\,|\operatorname{Re} m_{1n}(z)|) < 0,\ \operatorname{Im} F_n\Big(\operatorname{Re} m_{1n}(z) + \mathrm{i}C_1\frac{\eta\phi\bar{\sigma}_1}{E^2+\eta^2}\Big) > 0,$$

where by definition

$$\operatorname{Im} F_n(m_{1n}, z) = -\operatorname{Im} m_{1n} + \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i \operatorname{Im}(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2})}{\operatorname{Re}^2(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2}) + \operatorname{Im}^2(z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+m_{1n}\xi_j^2})}.$$

The proof of the above inequalities is similar to (S.2) using (S.1). We briefly discuss the proof of the first inequality in which case by a discussion similar to (S.4)

$$\begin{aligned}\operatorname{Im} F_n(m_{1n}, z) &= \eta \operatorname{Re} m_{1n} + \frac{\phi\bar{\sigma}_1\eta(1+E)}{E^2+\eta^2(1+E\times \mathrm{o}(1))^2} \\ &\leq -C_2\frac{\phi\bar{\sigma}_1\eta E}{(E^2+\eta^2)} + \frac{\phi\bar{\sigma}_1\eta(1+E)}{E^2(1+\eta^2\times \mathrm{o}(1))+\eta^2(1+2E\times \mathrm{o}(1))} < 0.\end{aligned}$$

This completes our proof.

For the second statement, from the first statement, we see that it is valid to write $m_{1n}(E)$. Since $\vartheta_1 \gg d_1$ holds with high probability (see (S.6)), it suffices to prove that for some constants $0 < C_2 < 1 < C_1$, when $|E-\vartheta_1| \leqslant Cd_1$,

$$m_{1n}(E) \in \left[-C_1\frac{\phi\bar{\sigma}_1}{E}, -C_2\frac{\phi\bar{\sigma}_1}{E}\right]. \tag{S.7}$$

Due to simplicity, we again only focus on the proof of the upper bound. According to Theorem 6.2 and (S.2), we shall have that $F_n(m_{1n}(E)) = 0$. Moreover, since $F_n(m_{1n}(\vartheta_1)) = 0$, to prove (S.7), it suffices to prove

$$F_n(-C_2\phi\bar{\sigma}_1/E) < 0,\ F_n(-C_1\phi\bar{\sigma}_1/E) > 0. \tag{S.8}$$

Due to similarity, we focus our discussion on the first inequality. By definition, we have that

$$\begin{aligned}F_n(-C_2\phi\bar{\sigma}_1/E) &= C_2\frac{\phi\bar{\sigma}_1}{E} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{E - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1-\xi_j^2(C_2\frac{\phi\bar{\sigma}_1}{E})}} \\ &= C_2\frac{\phi\bar{\sigma}_1}{E} - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{E(1-\frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{E-C_2\xi_j^2\phi\bar{\sigma}_1})}.\end{aligned}$$

By a discussion similar to (S.7), we further have that

$$F_n(-C_2\phi\bar{\sigma}_1/E) = C_2\frac{\phi\bar{\sigma}_1}{E} - \frac{1}{n}\sum_{i=1}^p\frac{\sigma_i}{E(1-\mathrm{o}(1))} = (C_2-1+\mathrm{o}(1))\frac{\phi\bar{\sigma}_1}{E} < 0. \tag{S.9}$$

This completes the proof of the first statement.

Finally, we prove the third statement using the first two statements. For $z \in \mathbf{D}_u$, by definition, we have that

$$\begin{aligned}m_{2n}(z) &= \frac{1}{n}\sum_{j=1}^n \frac{\xi_j^2}{-E-\mathrm{i}\eta-(E+\mathrm{i}\eta)\xi_j^2(\operatorname{Re} m_{1n}+\mathrm{i}\operatorname{Im} m_{1n})} \\ &= \frac{1}{n}\sum_{j=1}^n \frac{\xi_j^2\left[(-E-\xi_j^2(E\operatorname{Re} m_{1n}-\eta\operatorname{Im} m_{1n})+\mathrm{i}\eta+\mathrm{i}\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))\right]}{(-E-\xi_j^2(E\operatorname{Re} m_{1n}-\eta\operatorname{Im} m_{1n}))^2+(-\eta-\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))^2}.\end{aligned} \tag{S.10}$$

According to the results in the first two statements, the definition of $\mathbf{D}_u$ in (S.5) and the elementary relation that $|m_{1n}(z)| = \mathrm{O}(1)$, we find that for some constants $C_1, C_2 > 0$, when n is sufficiently large

$$|(-E-\xi_j^2(E\operatorname{Re} m_{1n}-\eta\operatorname{Im} m_{1n})+\mathrm{i}\eta+\mathrm{i}\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))| \leqslant C_1 E,$$

and

$$(-E-\xi_j^2(E\operatorname{Re} m_{1n}-\eta\operatorname{Im} m_{1n}))^2+(-\eta-\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))^2 \geqslant C_2(E+\xi_j^2)^2.$$

Together with a discussion similar to (S.7), we readily see that for some large constant $C>0$

$$|m_{2n}(z)| \leq \frac{C}{n}\sum_{j=1}^n \frac{E\xi_j^2}{(E+\xi_j^2)^2} \leq \frac{C}{n}\sum_{j=1}^n \frac{\xi_j^2}{E-\xi_j^2} = \mathrm{O}(e_2). \tag{S.11}$$

Then together with the definition of $m_n(z)$, we see that

$$|m_n(z)| \leq \frac{1}{p|z|}\sum_{i=1}^p \frac{1}{|1+\sigma_i m_{2n}(z)|} \leq \frac{1}{p|z|}\sum_{i=1}^p \frac{1}{1+\sigma_i |m_{2n}(z)|} \leq \frac{1}{|z|} = \mathrm{O}(E^{-1}).$$

To control the imaginary part, by (S.10), we can write

$$\operatorname{Im} m_{2n}(z) = \frac{1}{n}\sum_{j=1}^n \frac{\xi_j^2(\eta+\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))}{(-E-\xi_j^2(E\operatorname{Re} m_{1n}-\eta\operatorname{Im} m_{1n}))^2+(-\eta-\xi_j^2(E\operatorname{Im} m_{1n}+\eta\operatorname{Re} m_{1n}))^2}.$$

Combining with (S.1) and (S.6), we see that for some constant $C>0$

$$\operatorname{Im} m_{2n}(z) \leqslant \frac{C}{n}\sum_{j=1}^n \frac{\xi_j^2(\eta+\xi_j^2\eta)}{(E+\xi_j^2(\mathrm{O}(1)+\eta^2\times \mathrm{O}(E^{-1})))^2+(\eta+\xi_j^2\times \mathrm{O}(1)+\eta\times \mathrm{O}(E^{-1}))^2}$$

$$= \mathrm{O}\Big(\frac{1}{n}\sum_{j=1}^n \frac{\eta\xi_j^4}{\mathrm{O}(E^2)}\Big) = \mathrm{O}\big(\frac{\eta}{E}\big), \tag{S.12}$$

where in the last step we used (S.25). Moreover, using the definition of $m_n(z)$ in (6.2), we can write

$$\operatorname{Im} m_n(z) = \frac{1}{p}\sum_{i=1}^p \frac{\eta+\sigma_i\eta\operatorname{Re} m_{2n}+\sigma_i E\operatorname{Im} m_{2n}}{(E+\sigma_i E\operatorname{Re} m_{2n}-\sigma_i\eta\operatorname{Im} m_{2n})^2+(\eta+\sigma_i\eta\operatorname{Re} m_{2n}+\sigma_i E\operatorname{Im} m_{2n})^2}.$$

Together with (S.11) and (S.12), we can easily see that

$$\operatorname{Im} m_n(z) = \mathrm{O}\big(\frac{\eta}{E^2}\big).$$

This completes our proof. □

REMARK S.6.1. We may observe from the proof of Lemma S.1.2 that in many cases we can directly write $m_{1n}(\vartheta_1)$ without considering its imaginary part. For example, when (2.3) holds, for $z_0=\vartheta_1+\mathrm{i}\eta$, using (S.3) and (6.5), we see that

$$\begin{aligned}\lim_{\eta\downarrow 0}\operatorname{Im} m_{1n}(z_0) &= \lim_{\eta\downarrow 0}\frac{1}{n}\sum_i \frac{\sigma_i(\eta+\frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^4\operatorname{Im} m_{1n}(z_0)}{|1+\xi_j^2 m_{1n}(z_0)|^2})}{|z_0-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z_0)}|^2}\\
&= \lim_{\eta\downarrow 0}\frac{1}{n}\sum_i \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n \frac{\xi_j^4\operatorname{Im} m_{1n}(z_0)}{|1+\xi_j^2 m_{1n}(z_0)|^2}}{|\vartheta_1-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z_0)}|^2}\\
&= \Big(\frac{1}{n}\sum_i \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n \frac{(\xi_{(1)}^2+d_2)^2\xi_j^4}{|\xi_{(1)}^2+d_2-\xi_j^2|^2}}{|\vartheta_1-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{(\xi_{(1)}^2+d_2)\xi_j^2}{\xi_{(1)}^2+d_2-\xi_j^2}|^2}\Big)\times\lim_{\eta\downarrow 0}\operatorname{Im} m_{1n}(z_0).\end{aligned}$$

Then by a discussion similar to (S.12), using the fact that $\alpha \geqslant 2$, we find that for any $\vartheta_1 \gtrsim \xi_{(1)}^2$

$$\lim_{\eta\downarrow 0}\operatorname{Im} m_{1n}(z_0) = \mathrm{o}(1)\times\lim_{\eta\downarrow 0}\operatorname{Im} m_{1n}(z_0),$$

which holds true if and only if $\lim_{\eta\downarrow 0}\operatorname{Im} m_{1n}(z_0)=0$. This shows that $m_{1n}(\vartheta_1)$ is well-defined for $\vartheta_1 \gtrsim \xi_{(1)}^2$.

S.6.1.2. Proof of Lemma S.1.5

Proof. We first prove the results when conditionally. Let Ω be the event satisfying (c) of Definition S.1.10. According to Lemma S.1.12, we find that $\mathbb{P}(\Omega) = 1 - \mathrm{o}(1)$. Now we choose a realization $\{\xi_i^2\} \in \Omega$ so that the proofs of parts (a) and (b) are purely deterministic.

Proof of (a). We start with (S.13). According to (6.2), we have that

$$m_{1n}(z) = \frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i}{-z + \frac{\sigma_i}{n} \sum_{j=1}^{n} \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}}.$$

To characterize the bulk of the spectrum, we take the imaginary part on the both sides of the above equation and let $\eta \downarrow 0$ to obtain that

$$\text{(S.13)} \qquad \operatorname{Im} m_{1n}(z) = \frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i^2 \left(\frac{1}{n} \sum_j \frac{\xi_j^4 \operatorname{Im}(m_{1n})}{\operatorname{Re}^2(1+\xi_j^2 m_{1n}) + \xi_j^4 \operatorname{Im}^2(m_{1n})} \right)}{(E - \operatorname{Re}(\frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}}))^2 + \operatorname{Im}^2(\frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})}.$$

The above equation can be further rewritten as

$$\text{(S.14)} \qquad 0 = \operatorname{Im} m_{1n}(z)(1 - g(m_{1n}, E)),$$

where $g(m_{1n}, E)$ is denoted as

$$g(m_{1n}, E) := \frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i^2 \left(\frac{1}{n} \sum_j \frac{\xi_j^4}{\operatorname{Re}^2(1+\xi_j^2 m_{1n}) + \xi_j^4 \operatorname{Im}^2(m_{1n})} \right)}{(E - \operatorname{Re}(\frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}}))^2 + \operatorname{Im}^2(\frac{\sigma_i}{n} \sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}})}.$$

Similar to the arguments used in [52, 56], it is easy to see that for any fixed $\operatorname{Re} m_{1n} < -l^{-1}$ and E, $g(m_{1n}, E) \to 0$ when $|\operatorname{Im} m_{1n}| \to \infty$, and $g(m_{1n}, E) \to +\infty$ in order to satisfy (S.14) when $|\operatorname{Im} m_{1n}| \to 0$.

Therefore, by monotonicity, there exists a unique $\operatorname{Im} m_{1n} > 0$ such that (S.14) holds, which corresponds to the bulk of the spectrum.

Furthermore, for any fixed $\operatorname{Re} m_{1n} > -l^{-1}$ and fixed E so that Theorem 6.2 holds, we have that $g(m_{1n}, E)$ is monotone decreasing in terms of $|\operatorname{Im} m_{1n}|$.

Let E_+ be defined according to $m_{1n}(E_+) = -l^{-1}$. In view of (S.3) and (S.2), we have that

$$l^{-1} = \frac{1}{n} \sum_{i=1}^{p} \frac{\sigma_i}{E_+ - \sigma_i \widehat{\mathsf{s}}_2}.$$

Let $\widetilde{\mathsf{s}}_3$ be defined similarly as $\widehat{\mathsf{s}}_3$ in (S.10) by replacing $\widehat{L}_+$ with E_+. Based on the above arguments and definitions, it is easy to see that

$$\sup_{\operatorname{Re} m_{1n} \in (-l^{-1}, \infty)} g(m_{1n}, E) = g(-l^{-1}, E_+) = \phi \widetilde{\mathsf{s}}_3.$$

Assuming that $\phi \widetilde{\mathsf{s}}_3 < 1$, we conclude that (S.14) holds only if $\operatorname{Im} m_{1n}(z) = 0$ which corresponds to the outside part of the spectrum. This shows that

$m_{1n} = -l^{-1}$ is at the right edge of the spectrum and gives the expression of the end point $\widehat{L}_+$ as in (S.13). Therefore, $\widehat{L}_+ = E_+$ and $\widehat{\mathsf{s}}_3 = \widetilde{\mathsf{s}}_3$. This completes the proof.

Second, the proof of (S.14) follows from an argument similar to Lemma 8.4 of [56] utilizing the estimate (2.5), we omit the details.

Third, we prove (S.15). The closeness of s_k and $\widehat{\mathsf{s}}_k, k = 1, 2, 3, 4,$ follows from arguments similar to the last equation of (S.27). Now we proceed to the proof of the second equation. According to the proof of (S.13) and an analogous argument, we found that

$m_{1n}(\widehat{L}_+) = -l^{-1}$ and $m_{1n,c}(L_+) = -l^{-1}$. Together with the definitions of $\mathsf{s}_2, \widehat{\mathsf{s}}_2$, m_{2n} and $m_{2n,c}$, we find that

$$\mathsf{s}_2 = -m_{2n,c}(L_+)L_+, \ \widehat{\mathsf{s}}_2 = -m_{2n}(\widehat{L}_+)\widehat{L}_+. \tag{S.15}$$

Next, by (S.13) and an analogous argument for L_+ (see (3.6) and the proof of part II below), we have that

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^p \frac{-l\sigma_i}{(-L_+ + \sigma_i \mathsf{s}_2)} + \frac{1}{n}\sum_{i=1}^p \frac{l\sigma_i}{(-\widehat{L}_+ + \sigma_i \widehat{\mathsf{s}}_2)} \\ &= \frac{1}{n}\sum_i \frac{-l\sigma_i}{(-L_+ + \sigma_i \mathsf{s}_2)} + \frac{1}{n}\sum_i \frac{l\sigma_i}{(-L_+ + \sigma_i \widehat{\mathsf{s}}_2)} + \frac{1}{n}\sum_i \frac{l\sigma_i(\widehat{L}_+ - L_+)}{(-\widehat{L}_+ + \sigma_i \widehat{\mathsf{s}}_2)(-L_+ + \sigma_i \widehat{\mathsf{s}}_2)}. \end{aligned} \tag{S.16}$$

By first equation of (S.15), (S.15) and Assumption S.1.1, we can conclude our proof.

Fourth, we work with (S.16) and (S.17). Due to similarity, we only prove (S.16). According to (6.2), we have that

$$m_{1n}(z) = \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{-z + \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}}.$$

Consequently, it is easy to see that for $z = \widehat{L}_+ - \kappa + \mathrm{i}\eta \in \mathbf{D}_b$,

$$m_{1n}(\widehat{L}_+) - m_{1n}(z) = R_1(\widehat{L}_+ - z) + R_2(m_{1n}(\widehat{L}_+) - m_{1n}(z)), \tag{S.17}$$

where we denote

$$R_1 := \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i}{(-\widehat{L}_+ + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})(-z + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})}$$

$$R_2 := \frac{1}{n}\sum_{i=1}^p \frac{\frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(\widehat{L}_+))(1+\xi_j^2 m_{1n}(z))}}{(-\widehat{L}_+ + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(\widehat{L}_+)})(-z + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})}.$$

To study the terms R_1 and R_2, we will need the following control whose proof follows from equations (4.24)-(4.28) of [56]

$$\frac{1}{n}\sum_{j=1}^n \frac{\xi_j^4}{(1-\xi_j^2 l^{-1})(1+\xi_j^2 m_{1n}(z))} = \begin{cases} \mathrm{O}\left(\log n\right), & d \geqslant 2; \\ \mathrm{O}\left(|l^{-1} + m_{1n}(z)|^{d-2}\log n\right), & 1 < d \leqslant 2. \end{cases} \tag{S.18}$$

For the denominator of R_1, since $z \in \mathbf{D}_b$, by a discussion similar to (S.16), we find they are bounded from below so that $R_1 = \mathrm{O}(1)$. Furthermore, since $m_{1n}(\widehat{L}_+) = -l^{-1}$, by a straightforward calculation, using the definition of $\widehat{\mathsf{s}}_2$ in (S.10) and the control (S.18), we observe that

$$\begin{aligned} R_1 &= \widehat{\mathsf{s}}_4 - \frac{1}{n}\sum_{i=1}^p \frac{\sigma_i(z-\widehat{L}_+) + \frac{\sigma_i^2}{n}\sum_j \frac{\xi_j^4(m_{1n}(z)+l^{-1})}{(1-\xi_j^2 l^{-1})(1+\xi_j^2 m_{1n}(s))}}{(-\widehat{L}_+ + \sigma_i\widehat{\mathsf{s}}_2)^2(-z + \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})} \\ &= \widehat{\mathsf{s}}_4 + \mathrm{O}(|z-\widehat{L}_+|) + \mathrm{O}(|m_{1n}(z)+l^{-1}|^{\min\{d-1,1\}}\log n), \end{aligned} \tag{S.19}$$

where in the second step we again used an argument similar to (S.16). Similarly, for R_2, we find that

$$R_2 = \phi\widehat{\mathsf{s}}_3 + \mathrm{O}(|z-\widehat{L}_+|) + \mathrm{O}(|m_{1n}(z)+l^{-1}|^{\min\{d-1,1\}}\log n). \tag{S.20}$$

We next provide a useful deterministic control. Using a discussion similar to [52, Lemma A.4], we find from (6.2) that

(S.21)
$$\frac{1}{n}\sum_{i=1}^{p}\frac{\sigma_i^2\frac{1}{n}\sum_{j=1}^n\frac{\xi_j^4}{(1+\xi_j^2 m_{1n}(z))^2}}{|-z+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}=1-\frac{1}{n}\sum_i\frac{\sigma_i\eta/\operatorname{Im} m_{1n}(z)}{|-z+\frac{\sigma_i}{n}\sum_j\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}=1-\eta\frac{|m_{1n}(z)|^2}{\operatorname{Im} m_{1n}(z)}.$$

Since $\operatorname{Im} m_{1n}(z)>0$, this implies that

$$0\leqslant 1-\frac{1}{n}\sum_i\frac{\sigma_i\eta/\operatorname{Im} m_{1n}(z)}{|(-z+\frac{\sigma_i}{n}\sum_j\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})|^2}\leqslant 1.$$

Together with Cauchy-Schwarz inequality, we see that

$$|R_2|\leqslant(\phi\widehat{\mathsf{s}}_3)^{1/2}\left(1-\frac{1}{n}\sum_i\frac{\sigma_i\eta/\operatorname{Im} m_{1n}(z)}{|(-z+\frac{\sigma_i}{n}\sum_j\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)})|^2}\right)^{1/2}<1,$$

where we used the fact $\operatorname{Im} m_{1n}(z)>0$ and $\eta>0$. Using (S.17), we find that $m_{1n}(\widehat{L}_+)-m_{1n}(z)\asymp\widehat{L}_+-z$. Then we can conclude our proof using (S.17), (S.19) and (S.20).

Finally, we prove the controls for the imaginary parts. For (S.18), the discussion is similar to that of Lemma 4.5 of [52]. According to (S.2), we find see that

(S.22)
$$\begin{aligned}-m_{1n}(z)&=\frac{1}{n}\frac{\sigma_1}{z-\frac{\sigma_1}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}}+\frac{1}{n}\sum_{i=2}^p\frac{\sigma_i}{z-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}}\\&=\mathrm{O}(\frac{1}{n\eta})+\frac{1}{n}\sum_{i=2}^p\frac{\sigma_i}{z-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}},\end{aligned}$$

where in the step we the fact that $\operatorname{Im} m_{1n}(z)\geqslant 0$ and the trivial bound that

(S.23)
$$\frac{1}{n}\Big|\Big(z-\frac{\sigma_1}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}\Big)^{-1}\Big|\leqslant n^{-1}\Big(\eta+\frac{\sigma_1}{n}\sum_{j=1}^n\frac{\xi_j^4\operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|^2}\Big)^{-1}\leqslant(n\eta)^{-1}.$$

Taking the imaginary part on both sides of (S.22), we see that for some constant $0<c<1$,

$$\begin{aligned}\operatorname{Im} m_{1n}(z)&=\frac{1}{n}\sum_{i=2}^p\frac{\sigma_i(\eta+\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^4\operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|})}{|z-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}+\mathrm{O}(\frac{1}{n\eta})\\&=\frac{1}{n}\sum_{i=2}^p\frac{\sigma_i\eta}{|z-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}+\frac{1}{n}\sum_{i=2}^p\frac{\frac{\sigma_i^2}{n}\sum_{j=1}^n\frac{\xi_j^4\operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|}}{|z-\frac{\sigma_i}{n}\sum_{j=1}^n\frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}+\mathrm{O}(\frac{1}{n\eta})\\&=\mathrm{O}(\eta)+\mathrm{O}(\frac{1}{n\eta})+c\operatorname{Im} m_{1n}(z),\end{aligned}$$

where in the last step we used discussions similar to (S.25) and (S.28) below. This concludes the proof. Then we prove (S.20) and (S.19) following [56, Lemma 5.2]. Due to similarity, we focus our analysis on $m_{1n}(z)$ and discuss $m_n(z)$ briefly in the end. In what follows, for notational simplicity, without loss of generality, we assume that on Ω, $\xi_{(i)}^2=\xi_i^2$. In what follows, we identify $\eta \equiv \eta_0$ till the end of the proof of the lemma. According to (6.2), we find that

$$\operatorname{Im} m_{1n}(z) = \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i(\eta + \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^4 \operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|^2})}{|z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}$$

$$= \frac{1}{n}\sum_{i=1}^{p} \frac{\sigma_i \eta}{|z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2} + \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\frac{\xi_1^4 \operatorname{Im} m_{1n}(z)}{|1+\xi_1^2 m_{1n}(z)|^2}}{|z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2} + \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_{j=2}^n \frac{\xi_j^4 \operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|^2}}{|z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}$$

(S.24)

$$= \mathsf{L}_1 + \mathsf{L}_2 + \mathsf{L}_3.$$

For the denominator, by the results and arguments in Section S.2.2.1, we observe that when $z \in \mathbf{D}_b'$

$$z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)} = z - \frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)} + \frac{\sigma_i}{n}\frac{\xi_1^2}{1+\xi_1^2 m_{1n}(z)}$$

$$= z - \frac{\sigma_i}{n}\sum_{j=2}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n,c}(z)} + \mathrm{O}((n\eta)^{-1} + n^{-1/2-1/(d+1)}).$$

Together with Assumption S.1.1 and (S.27), we find that for some small constant $c' > 0$, when n is sufficiently large,

$$\left| z - \frac{\sigma_i}{n}\sum_{j=1}^n \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)} \right| \geqslant c'.$$

This implies that

(S.25) $$\mathsf{L}_1 \asymp \eta.$$

For L_2, on the one hand, when $|z - z_0| \geqslant C n^{-1/2+3\epsilon_d}$, by (S.17), we conclude that on Ω, for some constant $C > 0$

(S.26) $$|\mathsf{L}_2| \leqslant n^{-1/2-3\epsilon_d} \operatorname{Im} m_{1n}(z).$$

On the other hand, when $z = z_0$ so that $\operatorname{Re} m_{1n}(z) = -\xi_1^2$, we can rewrite L_2 as

(S.27) $$\mathsf{L}_2 = \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i}{n \operatorname{Im} m_{1n}(z)}}{|z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2}.$$

Next, for L_3, by (S.45), (S.27) and the results and arguments in Section S.2.2.1, using the trivial bound for L_2 that $|\mathsf{L}_2| = \mathrm{O}((n\eta)^{-1})$, we conclude that when $z \in \mathbf{D}_b'$, for some constant $0 < \mathfrak{c} < 1$

(S.28) $$|\mathsf{L}_3| \leqslant \mathfrak{c} \operatorname{Im} m_{1n}(z).$$

Consequently, we find that (S.20) follows from (S.24), (S.25), (S.26) and (S.28). Moreover, (S.19) follows from (S.24), (S.25), (S.27) and (S.28) by solving the associated quadratic equation. Finally, we mention that the results for $\operatorname{Im} m_n(z)$ essentially follows from (6.2) that

$$\operatorname{Im} m_n(z) = \frac{1}{n}\sum_{i=1}^{p} \frac{\eta}{|z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2} + \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{1}{n}\sum_j \frac{\xi_j^4 \operatorname{Im} m_{1n}(z)}{|1+\xi_j^2 m_{1n}(z)|^2}}{|z - \frac{\sigma_i}{n}\sum_j \frac{\xi_j^2}{1+\xi_j^2 m_{1n}(z)}|^2},$$

with the results for $\operatorname{Im} m_{1n}(z)$. This completes our proof.

□

Proof of Part (b). For (S.21), on the one hand, when when $d > 1$ and $\phi^{-1} < \widehat{\mathsf{s}}_3$, the result has been proved in (S.4). On the other hand, when $-1 < d \leqslant 1$, we employ the proof idea as in the proof of Lemma A.3 of [54] using a continuity argument. Recall (S.12). Denote

$$g(x,y) \equiv \frac{\partial F_n(x,y)}{\partial x} + 1 = \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^{n}\frac{\xi_j^4}{(1+x\xi_j^2)^2}}{(-y+\frac{\sigma_i}{n}\sum_{j=1}^{n}\frac{\xi_j^2}{1+x\xi_j^2})^2}.$$

From our assumption that $-1 < d \leqslant 1$ and (2.5), we find that there exist constants $C, C_0 > 0$ such that $\mathrm{d}F(x) \geq C(l-x)^d \geq C_0(l-x)$ for $x \in (0,l)$. Let D be a sufficiently large constant and choose a sufficiently small constant $0 < \epsilon < D^{-1}$, we have that when n is sufficiently large, there exists some constants $C_1, C_2, C_3 > 0$

$$\begin{aligned}
g(-(l+\epsilon)^{-1}, \widehat{L}_+) &= \frac{1}{n}\sum_{i=1}^{p} \frac{\frac{\sigma_i^2}{n}\sum_{j=1}^{n}\frac{(l+\epsilon)^2\xi_j^4}{(l+\epsilon-\xi_j^2)^2}}{\left(\widehat{L}_+ - \frac{\sigma_i}{n}\sum_{j=1}^{n}\frac{(l+\epsilon)\xi_j^2}{l+\epsilon-\xi_j^2}\right)^2} \\
&\geq \frac{C_1}{n}\sum_{i=1}^{p}\frac{\sigma_i^2\int_{l-(D-1)\epsilon}^{l}\frac{(l+\epsilon)^2x^2}{(l+\epsilon-x)^2}\mathrm{d}F(x)}{(\widehat{L}_+ - \sigma_i \mathrm{O}(1))^2} \geq \frac{1}{n}\sum_{i=1}^{p}\frac{C_2\int_{l-(D-1)\epsilon}^{l}\frac{(l-x)}{(l+\epsilon-x)^2}\mathrm{d}x}{(\widehat{L}_+ - \sigma_i \mathrm{O}(1))^2} \\
&= \frac{1}{n}\sum_{i=1}^{p}\frac{C_2\int_{\epsilon}^{D\epsilon}\frac{(t-\epsilon)}{t^2}\mathrm{d}t}{(\widehat{L}_+ - \sigma_i \mathrm{O}(1))^2} \geqslant C_3(\log D - 1 + \frac{1}{D}) > 1,
\end{aligned}$$

for sufficiently large $D > 0$. Similar arguments apply to $g(-(l-\epsilon), \widehat{L}_+)$. Consequently, by the continuity of $g(x,y)$, we obtain that $\partial F_n(-l^{-1}, \widehat{L}_+)/\partial x > 0$. Since $\partial F_n(m_{1n}(\widehat{L}_+), \widehat{L}_+)/\partial x = 0$, we can conclude that (S.4) still holds. That is, $m_{1n}(\widehat{L}_+) > -l^{-1}$. This finishes the proof of (S.21).

For (S.22) and (S.23), using (S.21), by a discussion similar to (S.14), we see that

$$\frac{\partial^2 F_n(m_{1n}(\widehat{L}_+), \widehat{L}_+)}{\partial x^2} \asymp 1. \tag{S.29}$$

Armed with this input, the square root behavior of ρ at $\widehat{L}_+$ can be obtained in the same way as Lemma A.1 of [54]. Due to similarity, we omit the details. This completes our proof of Part (b). □

Finally, it is easy to check that we can follow lines of the proofs of parts (a) and (b) to prove the unconditional results by replacing the related quantities verbatim. We omit further details. □

S.6.2. Control of some bad probability events: proof of Lemma S.1.12

In this subsection, we prove Lemma S.1.12 case by case. We first prove Case (a) in Definition S.1.10.

Proof of Case (a). First, the last statement of (S.25) follows directly from strong law of large number. In fact, the result holds almost surely.

Then, we prove the second statement of (S.25). For the upper bound, since $\{\xi_i^2\}$ are independent, we readily see that when n is sufficiently large, for some constant $C' > 0$,

$$\mathbb{P}(\xi_{(1)}^2 \leqslant Cn^{1/\alpha}\log n) = \Big(1 - \mathbb{P}(\xi^2 > Cn^{1/\alpha}\log n)\Big)^n \geq \left(1 - \frac{L(Cn^{1/\alpha}\log n)}{(Cn^{1/\alpha}\log n)^\alpha}\right)^n$$
$$\geq (1 - C'n^{-1}\log^{-\alpha} n)^n \asymp \exp\left(-1/(C'\log^\alpha n)\right) \asymp 1 - \mathrm{O}(\log^{-\alpha} n).$$

where in the second step we used the assumption (2.3) and in the third step we used the assumption that $L(\cdot)$ is a slowly varying function. This proves the upper bound. Similarly, for the lower bound, we can show that for some large constant $C > 0$

$$\mathbb{P}(\xi_{(1)}^2 \leq n^{1/\alpha}\log^{-1} n) = \mathrm{O}(\log n/n^C). \tag{S.30}$$

This concludes the proof of the second statement.

Next, we prove the first statement using the second one. Note that

$$\mathbb{P}(\xi_{(1)}^2 - \xi_{(2)}^2 < n^{1/\alpha}\log^{-1} n) = \mathbb{P}(\xi_{(1)}^2 < n^{1/\alpha}\log^{-1} n + \xi_{(2)}^2) = \mathbb{P}(\xi_{(1)}^2 < Cn^{1/\alpha}\log^{-1} n) = \mathrm{O}(\log n/n^C),$$

where the second and third steps we used the results of the second statement.

Then we justify the fourth statement. In what follows, without loss of generality, we assume that n^b is an integer. For $c > 1$ and $b > 1/2$, we notice that for some large constant $C > 0$

$$\begin{aligned}
&\mathbb{P}(\xi_{(1)}^2 - \xi_{(n^b)}^2 < c^{-1}n^{1/\alpha}\log^{-1} n) \leq \mathbb{P}(\xi_{(n^b)}^2 \geq (1-c^{-1})n^{1/\alpha}\log^{-1} n) \\
&= \sum_{k=n^b}^{n} \binom{n}{k} \Big[\mathbb{P}(\xi^2 \geq (1-c^{-1})n^{1/\alpha}\log^{-1} n)\Big]^k \Big[\mathbb{P}(\xi^2 \leq (1-c^{-1})n^{1/\alpha}\log^{-1} n)\Big]^{n-k} \\
&= \sum_{k=n^b}^{n} \binom{n}{k} \left(\frac{\log^\alpha n}{n}\right)^k \left(1 - \frac{\log^\alpha n}{n}\right)^{n-k} \\
&\leq \sum_{k=n^b}^{n} \left(\frac{en}{k}\right)^k \left(\frac{\log^\alpha n}{n}\right)^k \left(1 - \frac{\log^\alpha n}{n}\right)^{n-k} \\
&\leq \sum_{k=n^b}^{n} \left(\frac{e}{k}\right)^k \log^{\alpha k} n e^{-\log^\alpha n(1-k/n)} = \mathrm{O}(\log n/n^C),
\end{aligned} \tag{S.31}$$

where in the first step we used (S.30), in the third step we used (2.3) and in the fourth step we used Stirling's formula. This concludes the proof.

Finally, we proceed to the proof of the third statement. Define a sequence of intervals $I_k := \{Cn^{1/\alpha}\log^{-1} n + kn^\epsilon, Cn^{1/\alpha}\log^{-1} n + (k+1)n^\epsilon\}, k = [\![1, n^{1/\alpha-\epsilon}]\!]$. It is easy to see that if $\xi_{(i)}^2 - \xi_{(i+1)}^2 < n^\epsilon$ when $\xi_{(i)}^2, \xi_{(i+1)}^2 \in I_k$ for some k. Setting $\mathsf{p}_k := \mathbb{P}(\xi^2 \in I_k)$, we see that

$$\mathbb{P}(|j \in [\![1,n]\!] : \xi_j^2 \in I_k| = 0) = (1 - \mathsf{p}_k)^n, \quad \mathbb{P}(|j \in [\![1,n]\!] : \xi_j^2 \in I_k| = 1) = n\mathsf{p}_k(1-\mathsf{p}_k)^{n-1}.$$

We now provide an estimate for p_k using (2.3). Note that

$$\begin{aligned}
\mathsf{p}_k &= \mathbb{P}(Cn^{1/\alpha}\log^{-1} n + kn^\epsilon \leq \xi^2 \leq Cn^{1/\alpha}\log^{-1} n + (k+1)n^\epsilon) \\
&\leq \frac{1}{(Cn^{1/\alpha}\log^{-1} n + kn^\epsilon)^\alpha} - \frac{1}{(Cn^{1/\alpha}\log^{-1} n + (k+1)n^\epsilon)^\alpha}
\end{aligned}$$

$$\leq Cn^{-1}\log^{\alpha} n \frac{(1+(k+1)n^{-1/\alpha+\epsilon}\log n)^{\alpha}-(1+kn^{-1/\alpha+\epsilon}\log n)^{\alpha}}{(1+kn^{-1/\alpha+\epsilon}\log n)^{\alpha}}$$

$$\leq Cn^{-1}\log^{\alpha} n^{-1/\alpha+\epsilon}\log n = Cn^{-(1+1/\alpha)+\epsilon}\log^{\alpha} n. \tag{S.32}$$

Armed with the above estimate, we see that when n is sufficiently large,

$$\mathbb{P}(\xi^2_{(i)},\xi^2_{(i+1)}\in I_k)\leq \mathbb{P}(|j\in[\![1,n]\!]:\xi_j^2\in I_k|\geq 2)=1-(1-\mathsf{p}_k)^n-n\mathsf{p}_k(1-\mathsf{p}_k)^{n-1}\leq n^2\mathsf{p}_k^2.$$

Consequently, together with (S.32), we have that for some constant $C_1>0$

$$\mathbb{P}(\xi^2_{(i)}-\xi^2_{(i+1)}\leq n^{\epsilon})\leq \sum_{k=1}^{n^{1/\alpha-\epsilon}} n^2\mathsf{p}_k^2\leq Cn^{1/\alpha+2-\epsilon}n^{-(2+2/\alpha)+2\epsilon}\log^{\alpha} n = n^{-1/\alpha+\epsilon}\log^{\alpha} n=\mathrm{o}(1), \tag{S.33}$$

as long as $\epsilon<1/\alpha$. This finishes the proof of the third statement. □

Then we prove Case (b) of Definition S.1.10.

Proof of Case (b). Due to similarity and for notational simplicity, we focus on the case $\beta=1$. The general setting can be proved analogously and we omit the details.

We start with the second statement of (S.26). For the upper bound, for any $C>1$, following Markov inequality, we have that for some universal constant $C'>0$ when n is sufficiently large, by (2.4),

$$\mathbb{P}(\xi^2_{(1)}<C\log n)=(1-\mathbb{P}(\xi^2\geq C\log n))^n\geq\left(1-\frac{\mathbb{E}e^{t\xi^2}}{e^{tC\log n}}\right)^n$$
$$=\left(1-\frac{C'}{n^{tC}}\right)^n\asymp\exp(-1/n^{tC-1})\asymp 1-\mathrm{O}(n^{-(tC-1)}).$$

We can therefore conclude our proof using $t=1$. Similarly, we can prove the lower bound that for some large constant $C_1>1$

$$\mathbb{P}(\xi^2_{(1)}\leqslant C^{-1}\log n)=\mathrm{O}(n^{-C_1}).$$

This completes the proof of the first statement.

For the first statement, the discussion is similar to (S.32). The main difference is that the sequence of intervals are defined as $I_k:=\{C^{-1}\log n+k\times C^{-1}\log n, C^{-1}\log n+(k+1)\times C^{-1}\log n\}, k\in[\![1,(C-C^{-1})/C^{-1}]\!]$. By Chernoff bound, we can control $\mathsf{p}_k:=\mathbb{P}(\xi^2\in I_k)$ as follows

$$\begin{aligned}\mathsf{p}_k&=\mathbb{P}(C^{-1}\log n+k\times C^{-1}\log n\leq\xi^2\leq C^{-1}\log n+(k+1)\times C^{-1}\log n)\\&=\mathbb{P}(\xi^2\geq C^{-1}\log n+k\times C^{-1}\log n)-\mathbb{P}(\xi^2\geq C^{-1}\log n+(k+1)\times C^{-1}\log n)\\&\leq C'(n^{-t(k+1)C^{-1}}-\inf_{t'>0}n^{-t'(k+2)C^{-1})})\\&\leq C'n^{-tC^{-1}},\end{aligned}$$

where $C'>0$ is some universal constant and in the third step we used (2.4). Now we choose t so that $tC^{-1}>2$. Then by a discussion similar to (S.33), we have

$$\mathbb{P}(\xi^2_{(1)}-\xi^2_{(2)}\leq C^{-1}\log n)\leq n^2\mathsf{p}_k^2\leq C'n^{2-tC^{-1}}.$$

This completes the proof of the first statement.

Finally, the last statement follows directly from the strong law of large number. In fact, the result holds almost surely. □

Finally we prove Case (c) of Definition S.1.10.

Proof of Case (c). Note that the fourth statement holds trivially and surely.

For the first statement, under the assumption of (2.5), we see that the lower bound follows from that

$$\begin{aligned}\mathbb{P}(l-\xi_{(1)}^2 > n^{-1/(d+1)-\epsilon_d}) &= \big(1-\mathbb{P}(l-\xi_{(1)}^2 < n^{-1/(d+1)-\epsilon_d})\big)^n \\ &\geq (1-Cn^{-\epsilon_d(d+1)-1})^n \\ &\geq 1-Cn^{-\epsilon_d(d+1)}.\end{aligned}$$

Similarly, for the upper bound, we find that when n is sufficiently large, for some constant $C'>0$

$$\begin{aligned}\mathbb{P}(l-\xi_{(1)}^2 > n^{-1/(d+1)}\log n) &\leq n\big(1-\mathbb{P}(l-\xi^2 \leq n^{-1/(d+1)}\log n)\big)^{n-1} \\ &\leq n\big(1-C^{-1}n^{-1}\log^{d+1} n\big)^{n-1} \\ &\leq ne^{-C^{-1}\log^{d+1} n} \leqslant n^{-C'}.\end{aligned}$$

This completes the proof of the first statement.

For the third statement, we prove by contradiction, i.e., there exists some sequence $\mathsf{a}_n = \mathrm{o}(1)$, $l-\xi_{\lfloor bn\rfloor} \leqslant \mathsf{a}_n$ holds with high probability. In fact, by a discussion similar to (S.31) using (2.5), we have that as long as $c\equiv c_n > n/\mathsf{a}_n$,

$$\begin{aligned}\mathbb{P}(l-\xi_{(c)}^2 \leqslant \mathsf{a}_n) &= \mathbb{P}(\xi_{(c)}^2 \geqslant l-\mathsf{a}_n) \\ &= \sum_{k=c+1}^n \binom{n}{k}\mathbb{P}(\xi^2 > l-\mathsf{a}_n)^k\mathbb{P}(\xi^2 \leq l-\mathsf{a}_n)^{n-k} = \mathrm{O}(n^{-C}),\end{aligned}$$

for some constant $C>0$ when n is sufficiently large. This completes our proof for the third statement.

For the second statement, its discussion is similar to (S.32). In this case, we will define the partition of the intervals as $I_k = [l-(k+1)n^{-1/(d+1)-\epsilon_d}, l-kn^{-1/(d+1)-\epsilon_d}]$ for $k = [\![1, n^{\epsilon_d}\log n]\!]$. Analogous to the arguments of (S.32), we have that

$$\mathsf{p}_k = \mathbb{P}(\xi^2 \in I_k) \leq Cn^{-\epsilon_d}n^{-1/(d+1)}(n^{-1/(d+1)}\log n)^d = Cn^{-1-\epsilon_d}\log^d n.$$

Using the above control with (S.33), we readily obtain that

$$\mathbb{P}(\xi_{(1)}^2 - \xi_{(2)}^2 \leq n^{-1/(d+1)-\epsilon_d}) \leq n^2\mathsf{p}_k^2 \leq Cn^{-2\epsilon_d}\log^{2d} n.$$

This completes the proof of the second statement.

Finally, we proceed to the proof of the last statement. Denote the random variable τ_{ξ_i} as follows

$$\tau_{\xi_i^2} := \frac{\xi_i^2}{1+\xi_i^2 m_{1n,c}(z)} - \int \frac{t}{1+tm_{1n,c}(z)}\mathrm{d}F(t).$$

By definition $\mathbb{E}\tau_{\xi_i^2} = 0$. On the one hand, according to the discussion around (S.45), we find that

$$\frac{1}{n}\sum_{i=1}^p \frac{\sigma_i^2\int \frac{t^2}{|1+tm_{1n,c}(z)|^2}\mathrm{d}F(t)}{|z-\sigma_i\int\frac{t}{1+tm_{1n,c}(z)}\mathrm{d}F(t)|^2} < 1.$$

Together with Assumption S.1.1 and the continuity of $m_{2n,c}$, we can therefore conclude that for some constant $C_0 > 0$,

$$\int \frac{t^2}{|1+tm_{1n,c}(z)|^2} \mathrm{d}F(t) < C_0.$$

As a consequence, by Cauchy-Schwarz inequality, we find that for some constants $C_1, C_2 > 0$

$$\mathbb{E}|\tau_{\xi^2}|^2 \leq C_1 \int \frac{t^2}{|1+tm_{1n,c}(z)|^2} \mathrm{d}F(t) < C_2 < \infty.$$

Since $\tau_{\xi_i^2}, 1 \leqslant i \leqslant n$, are independent, we can conclude our proof using Markov inequality. □

S.6.3. Fluctuation averaging arguments: Proof of Lemma S.3.5

In this section, we prove the fluctuation averaging results in Lemma S.3.5 following the strategies of Section 6 of [56]. Fluctuation averaging is a common step in the proof of local laws for random matrix models, especially when the LSD has a square root decay behavior near the edge so that the entries of the resolvents can be controlled under some ansatz; see the monograph [36] for a review. However, in our setting, due to the lack of square root decay as in (S.14), many entries of the resovelents, even the off-diagonal ones can be large when $\eta \sim n^{-1/2}$. To address this issue, we will follow the strategies of [56] to focus on the resolvent fractions instead of the entries themselves; see the discussion above Sections 6.1 of [52, 56]. In what follows, due to similarity, we focus on the parts which deviate from [56, Section 6] the most.

Proof of Lemma S.3.5. In what follows, with loss of generality, we assume that $\xi_1^2 \geqslant \xi_2^2 \geqslant \cdots \geqslant \xi_n^2$.

We start with part (1). Recall (S.9). Using Theorem S.1.9 and Remark S.2.9, we have that

$$|m_2 - m_2^{(1)}| \leq \left| \frac{1}{n} \frac{\xi_1^2}{z(1+\xi_1^2 m_{1n} + \mathrm{O}_{\prec}((n\eta_0)^{-1}))} \right| + \left| \frac{1}{n} \sum_{i=2}^{p} \frac{\mathrm{O}_{\prec}((n\eta_0)^{-1})}{z(1+\xi_i^2 m_{1n} + \mathrm{O}_{\prec}((n\eta_0)^{-1}))(1+\xi_i^2 m_{1n} + \mathrm{O}_{\prec}((n\eta_0)^{-1}))} \right|.$$

For the first term on the right-hand side of the equation, it can be trivially bounded by $(n\eta_0)^{-1}$ by a discussion similar to (S.23) using (S.18). The second term can also be controlled by $(n\eta_0)^{-1}$ using a discussion similar to (S.25). The proves the first equation in (S.27). For the second equation, due to similarity, we focus on $|m_2 - m_2^{(i)}|$. Using (S.19), we have that

$$|m_2 - m_2^{(i)}| \leq \frac{|\mathcal{G}_{ii}|}{n} + \frac{1}{n} \sum_{j \neq i} |\mathcal{G}_{jj} - \mathcal{G}_{jj}^{(i)}|. \tag{S.34}$$

For $\mathcal{G}_{ii}$, by Lemma S.1.13, Theorem S.1.9 and the assumption that $z \in \mathbf{D}_b'$ in (S.12), we conclude that with high probability, for some constant $C > 0$

$$|\mathcal{G}_{ii}| = \frac{1}{|z(1+\xi_i^2 m_1^{(i)} + Z_i)|} \leq C n^{1/(d+1)+\epsilon_d}. \tag{S.35}$$

For $\mathcal{G}_{jj} - \mathcal{G}_{jj}^{(i)}$, by Lemma S.1.13, (S.28) and Lemma S.1.15,

$$|\mathcal{G}_{jj} - \mathcal{G}_{jj}^{(i)}| = |\frac{\mathcal{G}_{ij}\mathcal{G}_{ji}}{\mathcal{G}_{ii}}| \prec |\mathcal{G}_{ii}||\mathcal{G}_{jj}^{(i)}|^2 \frac{\operatorname{Im} m_1^{(ij)}}{n\eta_0} \prec \frac{n^{2\epsilon_d}}{n} |\mathcal{G}_{ii}||\mathcal{G}_{jj}^{(i)}|^2,$$

where in the last step we used (S.18) and Theorem S.1.9. Inserting all the above bounds back to (S.34) and use the trivial bound that $|\mathcal{G}_{jj}^{(i)}| \leqslant \eta_0$, we can conclude the proof. This completes the proof of part (1).

We now proceed to the proof of parts (2) and (3). Due to similarity, we focus on the details of part (2) and briefly mention how to prove (3) in the end. For simplicity, following the conventions in [52, 56], we denote the operator

$$P_i := \mathbf{1} - \mathbb{E}_i,$$

where $\mathbb{E}_i$ is the conditional expectation with respect to $\mathbf{y}_i$. Using Lemma S.1.13, we see that on Ω

$$\frac{1}{n}\sum_{i=2}^{n} P_i(\frac{1}{\mathcal{G}_{ii}}) = \frac{1}{n}\sum_{i=2}^{n} P_i(-z - z\mathbf{y}_i^* G^{(i)}(z)\mathbf{y}_i) = -\frac{z}{n}\sum_{i=2}^{n} Z_i. \tag{S.36}$$

Consequently, it suffices to show that

$$\left|\frac{1}{n}\sum_{i} P_i(\frac{1}{\mathcal{G}_{ii}})\right| \prec n^{-1/2-\frac{1}{2}(\frac{1}{2}-\frac{1}{d+1})+2\epsilon_d}.$$

By Chebyshev's inequality, it suffices to prove the following lemma.

LEMMA S.6.2. *Under the assumptions of Lemma S.3.5, for any $z \in \mathbf{D}_b'$ and fixed even number $M \in \mathbb{N}$, we have*

$$\mathbb{E}^X \left|\frac{1}{n}\sum_{i=2}^{n} P_i(\frac{1}{\mathcal{G}_{ii}(z)})\right|^M \prec n^{M(-1/2-\frac{1}{2}(\frac{1}{2}-\frac{1}{d+1})+2\epsilon_d)}.$$

PROOF. The proof strategy and technique follows closely from Section 6 of [56]. In what follows, we adopt the way how [52, Section 6.3] generalizes [56, Section 6.2] and only check the core estimates that have been used in [56]. We first provide some notations following the conventions of [56, Section 6.1]. For any subset $\mathcal{T}, \mathcal{T}' \subset \{1, \ldots, n\}$ with $i, j \notin \mathcal{T}$ and $j \notin \mathcal{T}'$, we set

$$F_{ij}^{(\mathcal{T},\mathcal{T}')} \equiv F_{ij}^{(\mathcal{T},\mathcal{T}')}(z) := \frac{\mathcal{G}_{ij}^{(\mathcal{T})}(z)}{\mathcal{G}_{jj}^{(\mathcal{T}')}(z)}.$$

In case $\mathcal{T} = \mathcal{T}' = \emptyset$, we simply write $F_{ij} = F_{ij}^{(\mathcal{T},\mathcal{T}')}$. With Lemma S.1.13, according to [56, Lemma 6.1], we have that for any subset $\mathcal{T}, \mathcal{T}' \subset \{1, \ldots, p\}$ with $i, j \notin \mathcal{T}$ and $j \notin \mathcal{T}'$, and $\gamma \notin \mathcal{T} \bigcup \mathcal{T}'$

$$F_{ij}^{(\mathcal{T},\mathcal{T}')} = F_{ij}^{(\mathcal{T}\gamma,\mathcal{T}')} + F_{i\gamma}^{(\mathcal{T},\mathcal{T}')} F_{\gamma j}^{(\mathcal{T},\mathcal{T}')},$$

and

$$F_{ij}^{(\mathcal{T},\mathcal{T}')} = F_{ij}^{(\mathcal{T},\mathcal{T}'\gamma)} - F_{ij}^{(\mathcal{T},\mathcal{T}'\gamma)} F_{j\gamma}^{(\mathcal{T},\mathcal{T}')} F_{\gamma j}^{(\mathcal{T},\mathcal{T}')}.$$

Moreover, we have that for $\gamma \notin \mathcal{T}$

$$\frac{1}{\mathcal{G}_{ii}^{(\mathcal{T})}} = \frac{1}{\mathcal{G}_{ii}^{(\mathcal{T}\gamma)}}\left(1 - F_{i\gamma}^{(\mathcal{T},\mathcal{T})} F_{\gamma i}^{(\mathcal{T},\mathcal{T})}\right).$$

In order to apply the techniques of [56, Section 6.2], we need to prove the following estimates

$$
\begin{aligned}
&|m_1(z)-m_{1n}(z)| \prec (n\eta_0)^{-1}, \quad \operatorname{Im} m_1(z) \prec (n\eta_0)^{-1}, \ \left|P_i(\frac{1}{\mathcal{G}_{ii}})\right| \prec (n\eta_0)^{-1}, \ i\neq 1, \\
&\max_{i\neq j} |F_{ij}(z)| \prec n^{-(1/2-1/(d+1))/2+\epsilon_d}, \quad i,j\neq 1, \\
&\max_{i\neq j} \left|\frac{F_{ij}^{(\emptyset,i)}(z)}{\mathcal{G}_{ii}(z)}\right| \prec (n\eta_0)^{-1}, \quad i,j\neq 1,
\end{aligned} \tag{S.37}
$$

First, the first part of (S.37) follows from Theorem S.1.9, Lemma S.2.8 and Remark S.2.9 (recall (S.36)). Second, for the second part of (S.37), by a discussion similar to (S.35), for $i\neq j$ and $i,j\neq 1$, we have that for some constant $C>0$, with high probability

$$
|\mathcal{G}_{ii}^{(j)}| \leqslant C n^{1/(d+1)+\epsilon_d}. \tag{S.38}
$$

Together with Lemma S.1.13, we see that for some constant $C>0$

$$
\begin{aligned}
|F_{ij}| &= |z\mathcal{G}_{ii}^{(j)} \mathbf{y}_i^* G^{(ij)} \mathbf{y}_j| \prec \left| z\mathcal{G}_{ii}^{(j)} \frac{1}{n} \|G^{(ij)}\Sigma\|_F \right| \leq C\left|\mathcal{G}_{ii}^{(j)} \Big(\frac{\operatorname{Im} m_1^{(ij)}}{n\eta}\Big)^{1/2}\right| \\
&\prec n^{1/(d+1)+\epsilon_d} \frac{1}{n\eta_0} = n^{1/(d+1)-1/2+2\epsilon_d},
\end{aligned}
$$

where in the second step we used (S.28) and in the third step we used (S.38) and the fact $z\in \mathbf{D}_b'$. Finally, for the third part of (S.37), using Lemma S.1.13, Lemma S.1.15 and (S.28), we see that

$$
\left|\frac{F_{ij}^{(\emptyset,i)}}{\mathcal{G}_{ii}}\right| = \left|\frac{\mathcal{G}_{ij}}{\mathcal{G}_{jj}^{(i)}\mathcal{G}_{ii}}\right| = \left|z\mathbf{y}_i^* G^{(ij)}\mathbf{y}_j\right| \prec \sqrt{\frac{\operatorname{Im} m_1^{(ij)}(z)}{n\eta}}.
$$

We can therefore conclude our proof using Lemma S.2.8, Remark S.2.9 and (S.29).

Using (S.37) and Assumption 2.1, we can follow the proof of Corollary 6.4 of [56] verbatim and conclude that for any $\mathcal{T},\mathcal{T}',\mathcal{T}''\in\{2,\dots,n\}$ with $|\mathcal{T}|,|\mathcal{T}'|,|\mathcal{T}''|\leq M$, where M is some large positive even integer, and for $z\in\mathbf{D}_b'$, we have that when $i\neq j, i,j\neq 1$,

$$
\begin{aligned}
&|F_{ij}^{(\mathcal{T},\mathcal{T}')}(z)| \prec n^{-(1/2-1/(d+1))/2+\epsilon_d}, \\
&\left|\frac{F_{ij}^{(\mathcal{T}',\mathcal{T}'')}(z)}{G_{ii}^{(\mathcal{T})}(z)}\right| \prec (n\eta_0)^{-1}, \quad \left|P_i\Big(\frac{1}{\mathcal{G}_{ii}^{(\mathcal{T})}}\Big)\right| \prec (n\eta_0)^{-1}.
\end{aligned} \tag{S.39}
$$

Once the key ingredients (S.37) and (S.39) have been proved, we can follow lines of [56, Lemma 6.6] or [52, Lemma 6.11] to conclude the proof. Due to similarity, we omit the details. □

This completes the proof of part (2). The proof of part (3) is similar except we need to following the proof of Lemma S.6.2 and [56, Lemma 6.12] to show

$$
\mathbb{E}^X\left|\frac{1}{n}\sum_{i=2}^n \frac{1}{(1+\xi_i^2 m_{1n}(z))^2} P_i(\frac{1}{\mathcal{G}_{ii}(z)})\right|^M \prec n^{M(-1/2-\frac{1}{2}(\frac{1}{2}-\frac{1}{d+1})+2\epsilon_d)}.
$$

We omit the proof and refer the readers to the proof of [56, Lemma 6.12] for more details. This completes the proof of Lemma S.3.5. □